DOCUMENT RESUME

ED 379 531 CE 068 481

AUTHOR	Mosenthal, Peter B.; Kirsch, Irwin S.
TITLE	Defining the Proficiency Standards of Adult Literacy in the U.S.: A Profile Approach.
PUB DATE	94
NOTE	127p.; Paper presented at the National Reading Conference (San Diego, CA (November 30-December 4, 1994).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC06 Plus Postage.
DESCRIPTORS	Adult Basic Education; *Adult Literacy; Educational Needs; Evaluation Methods; *Literacy Education; Measurement Techniques; *Needs Assessment; *Profiles; Standards; *Student Evaluation
IDENTIFIERS	Anthropometry; *Literacy Task Analysis

ABSTRACT

Research was conducted to illustrate how the profile approach to measurement could be used to define "standardized fits" between literacy task difficulty and adult proficiency. To provide understanding of this type of standard, the study began by comparing clothes anthropometry to educational measurement; in the former, the concern is with fitting clothing size to human size, whereas in the latter the concern is with fitting task difficulty with adult proficiency. To optimize such fits, researchers proposed that, just as a set of variables (for example, neck and sleeve size) and their constructs (length measured in inches) provides a common means for interpreting and relating clothes size to human size, there must be a similar means of interpreting and relating task difficulty to human proficiency. Some variables and their constructs can be identified and validated that characterize both task difficulty and adult proficiency on the prose, document, and quantitative scales of the Department of Labor's Workplace Literacy Assessment and the recent National Adult Literacy Survey. Similar to the anthropometric categories of "small," "medium," "large," and "extra-large" in clothing sizes, five levels of task difficulty and adult proficiency were described and validated on the two assessments' three literacy scales. The study concluded by considering how these five levels serve as useful standardized definitions of "growth space," which, in turn, provides an important basis for designing enhanced computer-based measurement and instructional systems. (Contains 72 references.) (Author/KC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 379 531

Defining the Proficiency Standards of Adult Literacy in the U. S.:

A Profile Approach

Peter B. Mosenthal
Syracuse University
170 Huntington Hall
Syracuse, NY 13244
Phone: (315) 443-4757
Fax: (315) 443-5732
PMOSE@SUED. SYR. EDU

Irwin S. Kirsch
Educational Testing Service
Adult Literacy Learning and Assessment Group
Princeton, NJ 08541
Phone: (609) 734-1516
Fax: (609) 734-1309

BEST COPY AVAILABLE

Running Head: Defining the Proficiency Standards of Adult Literacy in the U. S.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P B Mosenthal

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CE 068 451

ABSTRACT

The purpose of this paper was to illustrate how the "profile approach" to measurement (Kirsch & Mosenthal, 1991; Mosenthal & Kirsch, 1989) could be used to define "standardized fits" between literacy task difficulty and adult proficiency. To provide understanding of this type of standard, the paper begins by comparing clothes anthropometry to educational measurement; in the former, the concern is with fitting clothes size to human size, while in the latter the concern is with fitting task difficulty with adult proficiency. To optimize such fits, it is argued that, just as a set of variables (e.g., neck and sleeve size) and their constructs (e.g., length measured in inches) provides a common means for interpreting and relating clothes size to human size, there must be a similar means for interpreting and relating task difficulty to human proficiency. To this end, the paper identifies and provides validation of variables and their constructs which characterize both task difficulty and adult proficiency on the prose, document, and quantitative scales of the Department of Labor's Workplace Literacy Assessment (DOL) (Kirsch et al., 1993) and the recent National Adult Literacy Survey (NALS) (Kirsch et al., 1994). Similar to the anthropometric categories of "small," "medium," "large," and "extra-large" in clothes and human sizes, five Levels of task difficulty and adult proficiency are described and validated on the two assessments' three literacy scales. The paper concludes by considering how these five Levels serve as useful standardized definitions of "growth space" which, in turn, provides an important basis for designing enhanced computer-based measurement and instructional systems (Bunderson et al., 1989).

Defining the Standards of Adult Literacy Proficiency in the U. S.: A Profile Approach

> To make standardized products that would meet the customer's expectations, it was necessary to have standardized units of measurement. We have become so accustomed to knowing what we mean by a minute, a foot, or a gallon that we forget how recently these simple terms were given clear definition. . . As late as 1892, there were eight different 'authoritative' values for the U. S. gallon.
>
> Daniel J. Boorstin (1973), *The Americans: The Democratic Experience*

Just as there were many definitive measures of the U. S. gallon in 1892, there are many definitive measures of American's adult literacy proficiency in 1994 (Kibby, 1993; Kirsch & Jungeblut, 1992; Kirsch & Mosenthal, 1989). Many of these measures are derived using a norm-referenced approach (Nafziger, Thompson, Hiscox, & Owen, 1975; Stedman & Kaestle, 1987). In this approach, a large number of multiple-choice literacy tasks are designed based on materials found in school settings, and then administered to a national sample of learners. Mean grade-level test scores are then determined. Learners whose scores are at or within grade level are called "average readers." Learners whose scores are significantly above grade level are called "good readers." Finally, learners whose scores are significantly below grade level are called "poor readers." In short, in this approach, grade-level scores become the standard by which literacy proficiency is gauged.

While the use of grade-level scores makes it possible to estimate the percentage of various population groups performing at, above, and below grade level, this approach carries with it certain assumptions and limitations when applied to adults. First, grade-level scores typically are determined based on children and adolescents' performance on school-based, multiple-choice tasks. In contrast, research (Heath, 1980; Kirsch & Guthrie, 1984; Mikulecky, 1982; Sticht, 1978; Venezky, 1982) has shown that the literacy materials and tasks which adults generally encounter in various

every-day contexts are quite different from those associated with school-based standardized tests. Consequently, performance on these school-based measures are not good predictors of performance on literacy tasks associated with nonschool settings.

A second limitation with using grade-level scores as a standard for measuring and interpreting adult literacy proficiency is that they represent the average performance of students functioning within a particular school setting and, thus, reflect much more than simply reading achievement. Standards of adult performance on such a scale indeed tend to be quite different from that of school-age children. Just as fourth-graders scoring at an eleventh-grade level on a test of reading achievement perform very differently from tenth-or eleventh-graders scoring at this same level, so adults scoring on the eighth-grade level are very different from seventh-or eighth-graders demonstrating this level of achievement (Mosenthal & Kirsch, 1989a; Stedman & Kaestle, 1987).

A third limitation with using grade-level scores as a standard for measuring and interpreting adult proficiency is that tasks on standardized, school-based literacy tests generally are selected for inclusion on the basis of item statistics designed to yield scores that maximally differentiate among individuals. Such a procedure can result in reliable and valid tests for purposes of ranking and selection, but has proven less useful for purposes of instructional placement, diagnosis of specific strengths and weaknesses, or for the certification of particular adult competencies (Cross & Paris, 1987; Nitko, 1989).

This limitation, in part, reflects the fact that analyses are rarely, if ever, undertaken to determine specific factors contributing to task difficulty within and between grade levels. Despite this fact, the preceding purposes are the very ones for which standardized reading achievement tests have been employed in literacy programs for adults (Mosenthal & Kirsch, 1989a). Concerns such as these have led researchers to adopt a "criterion-referenced approach" to measuring literacy proficiency.

Representing the criterion-referenced approach are national performance surveys (e.g., such as those conducted by Northcutt, 1975; Harris and Associates, 1970; Educational Testing Service (i.e., Murphy, 1973) and the National Assessment of Educational Progress, 1972, 1976) which attempt to go beyond school-related reading tasks by including materials more like those which adults typically encounter at home, at work, or while functioning within their communities. In each of these surveys, nonschool materials are sampled and used to develop tasks which are field-tested and then administered to various national samples.

With the exception of the Adult Functional Reading Survey (Murphy, 1973) (which presents results solely in terms of the percentage of adults who respond correctly to each task), most national performance surveys employ an additive scoring model. In this model, evaluators sum across tasks which readers get correct to yield a single performance score. Next, a single cut-score is selected. Based on this cut-score, readers are said to perform at one of two levels: "literate" or "illiterate" (the former being at or above the cut-score, the latter being below the cut-score).

While the criterion-reference approach has some advantages over norm-referenced approach, it, too, is not without significant limitations. First, as with the norm-referenced approach, no attempt is made in the criterion-referenced approach to analyze tasks with respect to the cognitive processes which underlie task difficulty or which distinguish literate from illiterate readers. Unfortunately, without such information, one cannot assume that different assessment instruments used to evaluate program effectiveness, to measure learner proficiencies, or to develop instructional programs are, in fact, focusing on the same aspects of literacy (Kirsch & Guthrie, 1984; Mosenthal & Kirsch, 1989a; Nitko, 1989).

A second limitation with the criterion-referenced approach is that there is no well-established basis for choosing one score over another as being *the* standard cut-score for defining the levels "literate" and "illiterate." Without such a basis, attempts

to determine such a score remain arbitrary at best, spurious at worst (Jaeger, 1989). A related problem, of course, is that, without knowledge of the constructs underlying a given instrument, it is impossible to compare the proficiencies of adults on two or more criterion-referenced tests. Because different tests may reflect different levels of difficulty, and because different tests tend to employ different cut-score criteria, no basis exists for comparing adult proficiencies between two or more criterion-referenced instruments.

Finally, a third limitation is that, in the criterion-referenced approach, reading proficiency is treated as an "all or none" competence. As Kirsch and Jungeblut (1986) have argued, literacy is not like being pregnant—a case where one either is or is not literate! Rather, most (if not all) adults have some degree of literacy proficiency. What differs is the fact that readers with greater literacy proficiency have a higher probability of getting more difficult literacy tasks correct than people with lower literacy proficiency.

This point suggests that, to understand literacy proficiency, we need to understand the nature of those tasks which readers with known literacy proficiencies are likely to perform. Recognizing that tasks with similar processing characteristics tend to cluster within definable ranges of difficulty (Kirsch, Jungeblut, & Mosenthal, 1994), we can use these ranges to define construct-valid levels of literacy proficiency which, in turn, can be used as a basis for defining performance standards for readers with different literacy proficiencies. Such a "level" of understanding is made possible when we consider how literacy proficiency is defined in the recent Department of Labor's Workplace Literacy assessment (DOL) (Kirsch & Jungeblut, 1992) and the National Adult Literacy Survey (NALS) (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993).

In light of the preceding discussion, the general purpose of this study was to describe and illustrate the possibility of using the five Levels of prose, document, and quantitative literacy proficiency from the recent DOL and NALS literacy surveys

(Kirsch & Jungeblut, 1992; Kirsch et al., 1993) as a basis for defining the standards of adult literacy proficiency in the U. S. The utility of these Levels is that they have the advantages of both grade-and cut-score levels, some additional advantages, and none of the disadvantages. Consider these advantages in turn,

First, the DOL and NALS Levels are derived from tasks which adults commonly encounter in the course of their daily lives. Similarly, these levels are based on tasks whose materials tend to occur in both general and workplace settings. Moreover, unlike grade-and cut-score levels, the DOL and NALS Levels are based on tasks with known processing characteristics which enable us to explain why tasks within one Level are harder or easier to process than tasks in other Levels (Kirsch, Jungeblut, & Mosenthal, 1994; Kirsch & Mosenthal, 1989a).

Based on this knowledge, we can go beyond simply comparing the proficiencies of different readers and the difficulty of different tasks; we can further explain the nature of the strategies which readers, with known background characteristics and proficiency scores, bring to bear on tasks representing different Levels of proficiency.

More importantly, this knowledge allows us go one step further. By understanding the constructs which relate reader characteristics to task characteristics, we now have a psychologically-grounded basis for determining proficiency standards for different adult groups and individuals. Educators can then use these standards for making much more precise decisions of how to test and instruct readers so that these readers' proficiencies and opportunities to learn can optimally be enhanced. In addition, researchers have a much more precise basis for matching tasks to subjects in a way which maximizes treatment sensitivity in small-scale studies. Finally, policy makers have a much more informed basis for defining and prioritizing problems and for establishing realistic literacy goals at the national, state, and local levels (Mosenthal, in press).

Although the DOL and NALS Levels have been generally described in previous summaries of adult literacy proficiency (Kirsch & Jungeblut, 1992; Kirsch et al., 1993; Kirsch et al., 1994), there has been no study completed to date which specifically describes the variables and constructs underlying these Levels. Nor has such a study been published which systematically illustrates how these variables and constructs apply to benchmark tasks representative of the five Levels which comprise each of the DOL and NALS three scales (i.e., prose, document, and quantitative). Finally, no study has been undertaken which begins to address the utility of these Levels as standard frameworks that can be used to measure and interpret adult literacy proficiency in research, testing, and instruction.

Our paper is organized as follows. First, we draw an analogy between clothes anthropometry and educational measurement. Here we make the comparison between fitting clothes to people representing different sizes, and fitting tasks to readers representing different literacy proficiencies. Building on this analogy, we illustrate how the use of the DOL and NALS' three scales, as well as the five proficiency Levels within each, allow us to match reader proficiencies to task difficulty.

In the second section of the paper, we identify, describe, and illustrate the variables and constructs underlying the DOL and NALS prose, document, and quantitative scales, respectively. Here we provide statistical evidence for the five Levels underlying each of the three scales.

Finally, in part three, we conclude by summarizing the salient characteristics of each of the five Levels, for each of the three scales. We then discuss the implications of these Levels as standards for measuring and interpreting adult literacy proficiency in light of four levels of educational measurement.

Equating Proficiency with Tasks Difficulty: An Anthropometry Problem

Anthropometry and Educational Measurement

As Boorstin (1973) has argued, our American Standard of Living has always been closely tied to our ability to establish common measures and then, using these measures, identify generalizable patterns among individuals and groups. Based on these patterns, goods and services can be tailored effectively and efficiently to address individuals or groups with shared characteristics. Among the many examples Boorstin uses to support this point is the rise of the ready-made clothing industry. Before 1860, the belief was that each person's body was unique. Given this assumption, there was no reason to manufacture large quantities of clothing that would fit different wearers. If a person wanted a proper fit, they made their own clothes or employed a personal tailor.

With the outbreak of the Civil War, this produced a sudden demand for uniforms in great quantities. Using inches as a standard metric, the army discovered that certain body measurements tended to recur in combination with predictable regularity. For instance, if a man's waist measured 38 inches and his sleeve-length was 34 inches, then the man's chest size would be 42 and 44 inches a high percentage of the time . This simple discovery made it possible to manufacture well-fitting clothes for a large population. Simple though it was, the discovery was essential to the ready-made-clothing industry, for, without it, no store could be provided with a disposable stock which could readily and inexpensively fit customers representing a variety of body shapes and sizes.

Over time, patterns of variables further emerged which enabled clothes manufacturers to aggregate variables (e.g., neck size and sleeve length). Thus, instead of producing different shirts representing multiple combinations of neck and sleeve length (e.g., one set of shirts with a neck size of 15.5 and a sleeve length of 32, and another set of shirts with a neck size of 15 and sleeve length of 30), manufacturers

began to produce clothes including sizes "small," "medium," "large," "extra large," and "extra-extra large." In accomplishing this, they added greater tolerance to the neck size and used the average sleeve length based on mean shoulder and arm length. Although this reduced the overall fit between shirts manufactured and their wearers, this greatly reduced production problems and costs, further making clothing even more affordable.

Such discoveries and practices gave rise to "anthropometry," or the measurement of individuals with a "view to discovering those patterns of physical and mental characteristics which recur among people differing in such dimensions as age, gender, and economic income." The importance of anthropometry was that it spurred the increase and diffusion of information about standards of measurement in numerous fields. For instance, based on this science, Daniel Ryan, in 1880, published his book *Human Proportions in Growth: Being the Complete Measurements of the Human Body for Every Age and Size during the Years of Juvenile Growth.* In short, this book served as an invaluable scientific guide for standardizing measurements in juveniles, boys, and men's ready-made clothing. (Besides being useful for the design of clothes, manufacturers found Ryan's information helpful in countless other ways, such as in improving the design for school furniture and setting standard room and door dimensions in home and office construction).

As anthropometry matured, this significantly enhanced America's Standard of Living. Notes Boorstin (1973, p. 189), "By the early twentieth century . . . for the first time in history, it was possible for a man to walk into a clothing store, indicate that he was a '42' and put on a jacket that, with little or no alteration, would satisfy a fastidious eye. People thus began to think of themselves as belonging to certain 'sizes'--in shoes, shirts, trousers, and hats . . ." In short, through the science of anthropometry, individuals who previously never could afford new clothes now could afford them. Anthropometry, based upon its use of standardized measurement, made it possible that the benefit of a few could now be the benefit of many.

Although the label "anthropometry" is no longer used in education *per se*, its cousin "educational measurement" remains a widespread concern (Linn, 1989). As Bunderson, Inouye, and Olsen (1989, p. 368) have defined it, this type of measurement ". . . is the process of specifying the position, or positions, for educational purposes, of persons, situations, or events on educationally relevant scales under stipulated conditions." When we analyze this definition, we see that it is not all that different from the anthropometry of relating patterns of clothes sizes to characteristics of body dimensions. First, both require the development of standardized scales. This involves, in part, the selection of a common unit of measure. In clothes anthropometry, the unit of measure is usually the "inch," which can be used to measure the waist of a pair of pants and the waist of an individual. Similarly, in educational measurement, this unit often is a set of response probabilities (e.g., "percentage-correct values" for each task) which can be used as a measure of task difficulty as well as a measure of reader proficiency.

Second, both the anthropometry of clothes and educational measurement involve the specification of a position, or positions, along relevant scales. More simply put, this means that, over time, individuals or groups with unknown body part sizes will have to be measured in order to relate them to a scale comprised of, say, different pant sizes or tasks representing different levels of difficulty. Some of these individuals will have smaller inseam and waist sizes while others will have larger sizes; some individuals will have lower proficiencies while others will have higher proficiencies. From time to time, these same individuals or groups will need to have their positions recalibrated along these scale dimensions, as people may change in their weight and proficiency.

Finally, clothes anthropometry and educational measurement are similar to the extent that the objects of both are people with different patterns of characteristics. Both anthropometry and educational measurement must make sense of these patterns in effective and efficient ways. In the former, the critical concern is with the size and shape of people. In attempting to relate the size and shape of clothes to the size and shape

of people, numerous measures based on an extremely large sample have to be taken before a few critical variables can be identified which characterize the various physical dimensions of most people. In the latter, the critical concern often is with the underlying proficiencies of individuals or groups. In attempting to relate the difficulty of tasks required by society to the proficiencies of society's members, numerous measures based on a national sample have to be administered so that critical variables can be determined which link task difficulty with proficiency.

DOL and NALS as Examples of Educational Measurement

In educational measurement, the use of standardized measures to identify generalizable patterns among individuals and groups has been achieved, to a large degree, through the use of national assessments, such as the recent Department of Labor's Workplace Literacy assessment (Kirsch & Jungeblut, 1992) and the National Adult Literacy Survey (NALS) (Kirsch et al., 1993). These assessments build upon the previous adult literacy assessment of the NAEP Young Adult Literacy Survey (YALS) (Kirsch & Jungeblut, 1986). YALS profiled the literacy skills of young adults, ages 21 to 25, who resided in the continental United States between April through September, 1985.

In extending YALS, the DOL assessment surveyed, as its target population, all adults in the continental United States who, at the time of the assessment (November, 1989, through June, 1990), were eligible to enroll (or who were actually enrolled) in JTPA programs; had applied for jobs through the ES system; or had filed claims for UI benefits. Interviews were conducted with 5,778 individuals—2,501 from JTPA and 3,277 from the combined ES and UI subpopulations. The 2,501 JTPA interviews were completed with a sample which represented approximately 1,100,000 adults in the U. S. The 3,277 ES/UI interviews were completed with a sample which represented approximately 18,937,087 adults.

In order to make comparisons of the weighted sample of respondents with the weighted sample of registrants in each of the programs, each selected JTPA, ES, and UI office kept records of demographic information for all registrants on the sampling days. For the most part, the differences in the weighted frequencies for the respondents and for all the registrants were within the bounds to be expected given sampling variability. (For additional details and considerations involving the sampling, weighting, and data collection activities, see Kirsch and Jungeblut, 1992.)

In further extending the DOL assessment, NALS took as its target population, all adults in the continental United States who, at the time of the assessment (i.e., the first eight months of 1992), were 16 years of age and older. The 26,091 interviews were completed with a sample which represented approximately 191,289,000 adults in the U. S. In addition to over sampling Blacks and Hispanics, this survey sampled the prison population. (For additional details and considerations involving the sampling, weighting, and data collection activities, see Kirsch et al., 1993.)

Like clothes anthropometry which has a common metric (i.e., the "inch"), the common unit of measure on the DOL and NALS was a weighted RP80 score; this score represents an estimate that a person will respond correctly to a particular task from a pool of tasks with an 80 percent probability. This probability is given as a function of a single parameter characterizing the proficiency of that person and one or more parameters characterizing the properties of the task (Hambleton, 1989).

The particular IRT model employed in ETS' adult surveys was the three-parameter logistic model. In this model, the task parameters included "task discrimination," "task difficulty," and "lower asymptote." *Task discrimination* is the rate of change in the probability of obtaining the correct response to a given item in relation to the reader's proficiency. *Task difficulty* is the general level of difficulty of a given item. The *lower asymptote* is the coefficient indicating the probability of a correct response by readers with very low proficiency. In addition to using RP80

values, a second unit of measure on the DOL and NALS was the Level of literacy proficiency. This unit of measure reflects the common processing characteristics of tasks within ranges of different difficulty levels. As such, tasks in lower levels tend to share characteristics which render them easy to process, while tasks at higher levels tend to share characteristics which render them more difficult to process.

Also like clothes anthropometry, DOL and NALS attempted to specify the position of people (as well as tasks) along several scales. On the one hand, literacy proficiencies of people were described in terms of RP80 values which ranged from 0 to 500. Using item response theory (or IRT) scaling procedures, the relative position of tasks on three literacy scales was defined in terms of a response probability of 80 percent (i.e., RP80). For example, given a group mean of 300, an individual in that group who scores at this mean could be expected to perform tasks at the 300 level with an 80 percent probability of success.

On the other hand, the literacy proficiencies of people were also described in terms of Level scores which represent five levels of task difficulty. Level 1 included tasks whose RP80 values are below 225. Level 2 included tasks whose RP80 values range from 225 to 275. Level 3 included tasks whose RP80 values range from 276 to 325. Level 4 included tasks whose RP80 values range from 326 to 375. Finally, Level 5 included tasks whose RP80 values are above 375. Reporting people's proficiencies in terms of Level scores is not unlike reporting people sizes in terms of "small," "medium," "large," "extra large," and "extra-extra large." In short, these Levels represent a categorical scale which builds upon the continuous RP80 scale ranging from zero to 500.

Based on both RP80s and Level, three scales were used in the DOL and NALS to characterize the different dimensions of literacy found in non-school settings. These scales included "prose," "document," and "quantitative" literacy. *Prose literacy* involves the knowledge and skills needed to understand and use information from texts

such as editorials, news stories, brochures, pamphlets, poems, and fiction; e.g., finding a piece of information in a newspaper article, interpreting instructions from a warranty, inferring a theme from a poem, or contrasting views expressed in an editorial.

Document literacy includes the knowledge and skills required to locate and use information contained in materials consisting of such things as job applications, payroll forms, transportation schedules, maps, tables, and graphs; e.g., locating a particular intersection on a street map, using a schedule to choose the appropriate bus, or entering information on an application form.

Lastly, *quantitative literacy* includes the knowledge and skills required to apply arithmetic operations, either alone or sequentially, using numbers embedded in printed materials; e.g., balancing a checkbook, figuring out a tip, completing an order form, or determining the amount of interest from a loan advertisement. The usefulness of these scales is that, like the distinction of sleeve length, waist size, and neck size, they collectively define the dimensions of their domain (i.e., literacy) more accurately than would a single scale, thus allowing for a more precise fit between tasks and proficiencies.

The advantage of using item response theory to characterize performance on the prose, document, and quantitative literacy scales is that, given tasks with a known RP80 value and Level score, one can accurately predict how readers, representing different levels of literacy proficiency, are likely to complete these tasks. For example, consider someone who is estimated to be performing at 250 on the prose scale, as portrayed in Table 1. The information in the table shows that such an individual can be expected to perform tasks at this point with an 80 percent probability. In other words, such an individual would be expected to respond successfully to this task and others like it in a consistent manner eight out of ten times. An individual estimated to be performing at 250 on the scale has an 82 percent chance of responding correctly to the 246-level task

involving a magazine article. In addition, Table 1 shows that this individual would have even higher probabilities of success performing easier tasks, e.g., this person would have a 94 percent probability of success performing the 209-level task listed in Table 1. On the other hand, this same individual could be expected to respond to tasks near the 300 level with a probability of around 40 percent.

Insert Table 1 about here.

While Table 1 shows how readers are likely to perform given their known proficiencies, this table also illustrates how the construct of an RP80 operationally defines "task difficulty." In short, note that, as each document task's RP80 value increases, the likelihood of people being able to perform generally decreases. For instance, in Table 1, we observe that tasks with RP80 values around 210 have, on average, a higher probability of being processed than do tasks with RP80 values around 253. In turn, tasks with RP80 values around 253 have, on average, a much higher probability of being processed than do tasks with RP80 values around 346.

Similar equating between task difficulty and reader proficiency can be done using Level scores, as shown in Figure 1. This graph displays the probability that individuals performing at selected points on the prose, document, and quantitative scales will give a correct response to tasks with varying difficulty values. We see, for example, that a person with prose proficiency at 150 has a 50 percent chance of responding correctly to tasks in Level 1. Individuals with prose proficiency at 200, on the other hand, have an almost 80 percent probability of responding correctly to such tasks.

Insert Figure 1 about here.

Also similar to clothes anthropometry, the DOL and NALS examined people representing different patterns of characteristics in relation to a set of concomitant considerations. In clothes anthropometry, characteristics of body part sizes are studied

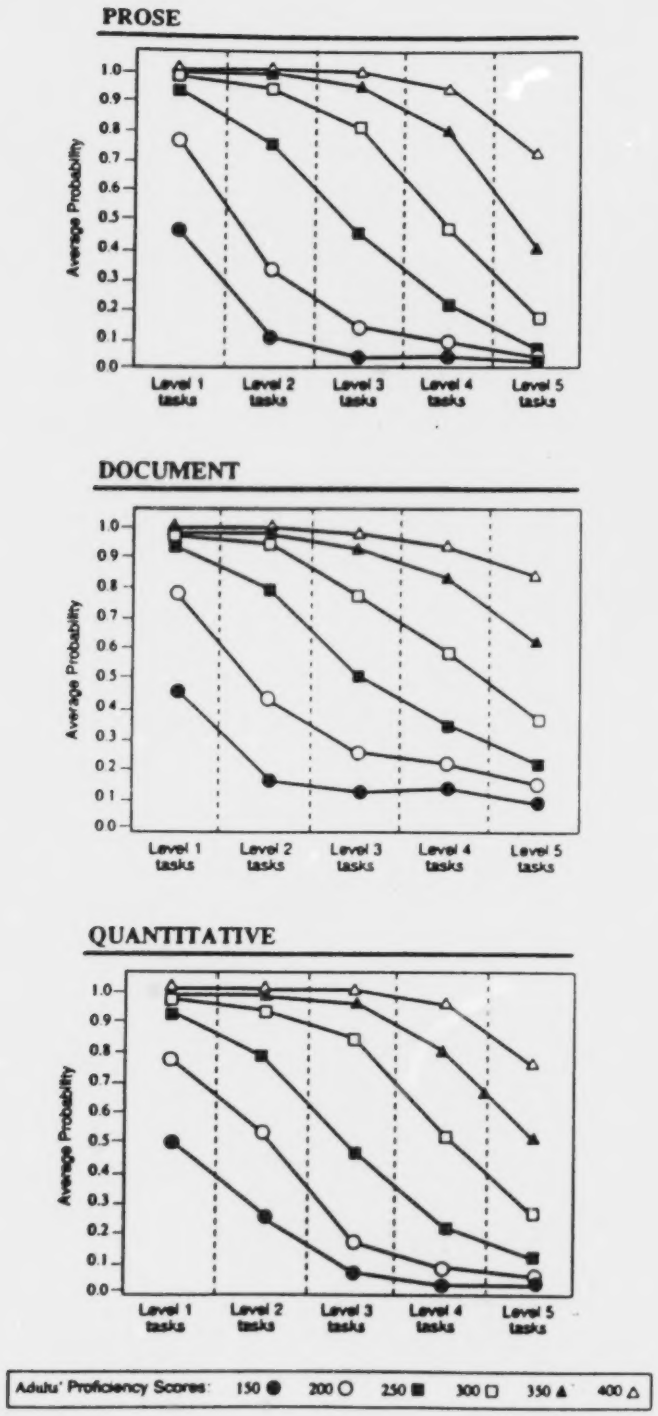


Figure 1. Average probabilities of successful performance by individuals with selected proficiency scores on prose, document, and quantitative tasks in each of the five literacy levels.

BEST COPY AVAILABLE

Table 1

Selected Prose Tasks with their RP80 Values and Probabilities of Performance by Adults Representing Different Proficiencies

Descriptions of Selected Tasks	RP80 Value	Associated Probabilities at Selected Proficiency Levels						
		150	200	250	300	350	400	450
Identify single piece of information in a brief sports article	209	36	75	94	99	100	100	100
Identify a single piece of information in a short announcement	210	40	75	93	98	100	100	100
Locate information in lengthy magazine article	246	11	43	82	97	99	100	100
Match two features of information in a brief sports article	253	13	42	78	95	99	100	100
Rephrase information stated in a magazine article	298	1	7	36	82	97	100	100
Integrate information from a news article on the economy	305	4	15	44	78	94	99	100
Compare new and old ways of processing credit card charges	346	3	10	28	57	82	94	98
Identify two situations that satisfy a given criterion	356	2	7	21	49	77	92	98

in relation to garment sizes. In DOL and NALS, the characteristics of proficiencies were studied in relation to literacy task complexity. In both cases, bridging each of these relations are a set of specified "constructs."

Constructs are variables and their interpretive statements which relate measures of a scale to applications of this scale (Hempel, 1966; Mosenthal, 1976-77). In terms of clothes anthropometry, the central variable is "length," which is used to define the size of clothes in terms of "sleeve length," "neck size," "length of inseam," and "waist size." Using the metric of inches, the measure of shirt sleeve length can be applied uniformly to the measure of arm and shoulder length. Since both sleeve, arm, and shoulder size are specified in terms of the same construct (i.e., "length"), this ensures a fit between the two with a high degree of probability.

In terms of DOL and NALS, there are two central variables which underlie task difficulty; these variables are *strategies* and *materials*. Included among strategies are the variables "type of match," "plausibility of distractors," "type of information," "operation specificity," "type of calculation," and "structural complexity." Included among materials are not length of inseam and waist size but rather the variables of "number of syllables per 100 words of text," "number of sentences per 100 words of text," "readability," "number of labels," number of items," and "type of document." Using the metrics RP80 and Level scores, the concept of task difficulty can be related rather uniformly to reader proficiency based on strategy and material variables.

In the next section, we discuss and illustrate how these strategy and material variables contribute to task difficulty on the DOL and NALS prose, document, and quantitative literacy scales.

Prose Literacy

An important area of literacy is the knowledge and skills needed to understand and use information organized in sentence and paragraph formats. Given the range of text types organized in such formats, the DOL and NALS assessments used prose materials that were primarily expository (i.e., materials which describe one or more states or an actions) (Mosenthal & Kirsch, 1991b) since such materials constitute much of the prose which adults read (Kirsch & Jungeblut, 1986; Kirsch et al., 1992). In addition, some narrative texts and poetry were included. The prose materials were drawn from newspapers, magazines, books, brochures, and pamphlets, and were reprinted in their entirety, using the typography and layout of the original source. As a result, the materials varied widely in length, density of information, and the use of structural or organization aids, such as section or paragraph headings, italic or bold face type, and bullets.

Prose Variables

Prose tasks involve the problem of first identifying "given" and "requested" information (Clark & Haviland, 1977; Kirsch & Mosenthal, 1992). *Given information* is information which is known and assumed to be true based on the way a question or directive is stated. *Requested information* of a question or directive is information which is being sought. To illustrate this, consider the question, "In the past five years, how many times has Susan Butcher won the Iditarod Sled Dog Race?" The given information in this instance is "In the past five years, Susan Butcher won the Iditarod Sled Dog Race one or more times." The requested information of this sentence is "*How many times* did Susan Butcher win?"

In processing prose, tasks tend to be easy when the requested information is concrete; tasks tend to be more difficult the more abstract the requested information

becomes. Hence, a task whose requested information involves a person or thing (e.g., a "who" or "what" question) tends to be easier to answer than a task whose requested information asks for a reason, purpose, or cause (e.g., "why"). In the DOL and NALS assessments, we referred to these different degrees of abstractness of requested information as "type of information."

Another dimension of prose processing requires readers to match information in a question or directive to corresponding information in a text. This involves the strategies of "locating," "cycling," "integrating," and "generating information." *Locating* involves the process of matching information based on given and/or requested information in a question to corresponding information in text. *Cycling* involves the process of making several locate matches within or between paragraphs to identify two or more pieces of information which relate to a common set of conditions (e.g., each piece of information represents a reason for not spanking children). *Integrating* involves the process of comparing or contrasting information once it has been identified via cycling. Finally, *generating* requires readers to use special background knowledge to relate information in a question or directive to information in text, or to select one plausible answer over another. In the DOL and NALS, these processes were represented by the variable "type of match" (Kirsch et al., 1994).

Yet another dimension of prose processing involves the situation where information in a text meets some but not all the conditions of information requested in a question or directive. In short, this information represents "distractors." The more conditions that distractors shares with an answer and the more closely distractors are positioned near (or next to) the correct answer, the more plausible the distractors are and the more difficult processing becomes. We referred to this variable as "plausibility of distractors" (Kirsch et al., 1994).

In addition to the three process variables, we considered a fourth variable representing material complexity. This variable was based on Fry's (1977) readability

formula and includes the average number of syllables per 100 words, the average number of sentences per 100 words, and readability level. The purpose for considering this variable was to provide a general description of the range of difficulty which the materials on the two assessments represented.

These four variables are described in detail below. How these variables relate to the five Levels of prose task difficulty is discussed. These variables are next illustrated using a variety of prose tasks from the DOL and NALS assessments, whose characteristics, in part, are shown in Table 2 below. The statistical significance of these variables as predictors of prose task difficulty is then presented.

Structural complexity. To analyze the structural complexity of prose used in the DOL and NALS assessments, we first counted the number of words in each of the 22 stimuli. We next divided each text into 100-word units. In those instances where a stimulus had fewer than 100 words, the number of words in that passage were divided into a 100; the resulting figure was then used as a multiplier to equate the number of syllables and sentences to the Fry (1977) readability graph. For example, if a stimulus consisted of 50 words, 71 syllables, and four sentences, "50" was then divided into "100 words," which produced "two" as the equating factor to use Fry's readability graph. Both "71 syllables" and "four sentences" were multiplied by "two" to determine readability level.

Each 100-word unit in each stimulus was then analyzed in terms of the number of syllables and sentences. Using these results, Fry's readability graph was consulted to determine the readability of each passage. As shown in Table 2, readability across the prose stimuli on the DOL and NALS scales ranged from fourth-to 15th grade level. The mean readability level was eighth grade (with a standard deviation of 2.25) and a median of eight. The mean values for readability by Level were 6.67 for Level 1; 7.41 for Level 2; 8.04 for Level 3; 8.88 for Level 4; and 9.00 for Level 5.

Insert Table 2 about here.

Process variables. The second set of variables which has been shown to influence prose task difficulty include three process variables (Kirsch et al., 1994): "Type of information," "type of match," and "plausibility of distractors." *Type of information* refers to the kind of information which users must identify to complete a question or directive. As Kirsch et al. (1994) have noted, prose questions generally consist of a rather restricted range of information types. These information types form a continuum of concreteness which was operationalized as follows.

Questions requesting information regarding the identification of *persons, animals, places* (as a noun), and *things* (e.g., "Underline the sentence that tells what Ms. Chanin ate during the swim?" (answer: 'banana and honey sandwiches, hot chocolate, lots of water and granola bars')) were scored the highest (i.e., **1**) in terms of concreteness.

Questions requesting information regarding the identification of *amounts, times, attributes, types, actions,* and *locations* (e.g., "In what state is the Toyota FX 16 built?" (answer: 'California')) were assigned a concreteness score of **2**.

Questions requesting information regarding the identification of *manner, goal, purpose, alternative, attempt, condition, pronominal reference,* and *predicate adjectives* (e.g., "Underline the sentence that tells how the Social Security Administration defines the term 'blind.'?" (answer: '*Blind* means the vision in your better eye is 20/200 or less or you have a limited visual field of 20 degrees or less')) were assigned a concreteness score of **3**.

Questions requesting information regarding the identification of *cause, effect, reason, result, evidence, similarity,* and *explanation* (e.g., "Underline the sentence that tells why Terry went home after visiting the United States?" (answer: 'I decided to go

Table 2

Characteristics of Variables (i.e., Readability ('Read'), Type of Match ('TOM'), Plausibility of Distractors ('POD') and Type of Information ('TOI')) for DOL and NALS Prose Tasks by RP80 and Difficulty Level

Description	RP80	Level	Read	TOM	POD	TOI
My Dream	149	1	4	1	1	1
Toyota, Acura, Nissan	182	1	8	1	1	1
Swimmer Marathon*	207	1	8	1	1	1
Blood Donor Pamphlet	210	1	7	1	1	2
Butcher Captures Iditarod	210	1	9	1	1	2
My Dream	224	1	4	1	2	4
Are You Eligible for SSI?	226	2	6	1	1	3
Summons for Jury Service	233	2	7	3	2	2
Growing Up	238	2	8	3	2	1
Blood Donor Pamphlet	241	2	7	1	2	3
PHP Subscriber Letter	246	2	10	3	1	3
Toyota, Acura, Nissan	246	2	8	3	2	4
Dr. Spock Column	247	2	8	2	2	3
Swimmer Marathon*	250	2	8	3	4	2
Getting More Miles per Gallon	253	2	10	3	2	1
Getting More Miles per Gallon	256	2	10	3	2	2
Shadows Columbus Saw	258	2	9	3	1	2
High Blood Pressure	261	2	7	3	2	4
Illegal Questions	262	2	6	3	2	2
Without Benefit*	262	2	4	3	1	3
$150,000 to Raise a Kid	274	2	6	2	4	2
Capital Gains	275	2	7	4	1	3
Returning Appliances*	275	2	5	3	2	3
Questions for New Jurors	276	3	6	4	2	1
Rules for Financial Security	277	3	8	3	2	4
Are You Eligible for SSI?	277	3	6	4	2	3
Shadows Columbus Saw	279	3	9	3	4	1
Credit Card Bill Reply Letter*	280	3	7	3	2	4
Financial Security Tips	280	3	8	2	2	4
Dr. Spock Column	283	3	8	3	2	4
Growing Up	283	3	8	4	3	2
Growing Up	287	3	8	5	1	4
Se Habla Espanol Hits Chicago	288	3	15	3	4	3
Food and Nutrition*	289	3	7	3	4	1
PHP Subscriber Letter	292	3	10	4	3	2
Summons for Jury Service	296	3	7	4	2	4
Shadows Columbus Saw	298	3	9	3	2	3
Shadows Columbus Saw	298	3	9	3	2	2
Summons for Jury Service	301	3	7	4	3	3
Economic Index Slip Is Hailed*	304	3	10	3	2	4
Dr. Spock Column	311	3	8	3	3	4
Questions for New Jurors	313	3	6	4	2	3
Illegal Questions	315	3	6	4	3	3
Ida Chen	315	3	7	4	2	3
Ida Chen	317	3	7	4	3	2
Toyota, Acura, Nissan	319	3	8	4	3	2
Fueled	324	3	9	5	1	3
Tom Wicker Article*	328	4	10	4	4	4
Shadows Columbus Saw	331	4	9	5	4	1
Financial Security Tips	332	4	8	5	5	2
Technology Creates Need*	333	4	8	4	4	4
Sloppy Work Perils Plant*	342	4	13	4	4	4
Handling Receipts	346	4	8	4	2	4
Fueled	346	4	9	5	1	4
XYZ's Benefits Chart*	347	4	9	6	2	5
Se Habla Espanol Hits Chicago	349	4	15	5	5	4

BEST COPY AVAILABLE

New Method for Capital Gains	356	4	7	5	5	3
Blood Donor Pamphlet	359	4	7	4	5	2
Getting More Miles per Gal.	359	4	10	6	2	5
Are You Eligible for SSI?	362	4	6	4	5	2
Dickinson's Honey Poem*	362	4	6	5	2	4
Technology Creates Need*	370	4	8	6	4	4
Fueled	374	4	9	6	2	4
Growing Up	383	5	8	6	2	5
New Method for Capital Gains	385	5	7	2	4	2
Toyota, Acura, Nissan	406	5	8	5	4	5
Questions for New Jurors	410	5	6	6	2	5
Ida Chen	423	5	7	6	3	5
Se Habla Espanol Hits Chicago	433	5	15	5	5	3
Handling Receipts	446	5	8	7	5	5
Sloppy Work Perils Plant*	468	5	13	6	5	4
*Tasks which appear in both DOL and NALS assessments.						

BEST COPY AVAILABLE

home and get a job to make more money to come back to the States')) were assigned a concreteness score of 4.

And finally, questions requesting information regarding the identification of *equivalent, difference, theme* (or *pattern*) were assigned a concreteness score of 5 ("equivalence," in this case, tended to be a highly unfamiliar term for which readers had to provide a definition from prior knowledge). An example of a question associated with "difference" information was "Identify and list two differences between the new and old ways American Express handles charge-card receipts" (ans. '(1) In the old way, information from paper receipts are microfilmed, while in the new way this information is transformed by image processing camera into electronic digital image; (2) in the old way, charge amounts are entered by computer operator from receipts, while in the new way charge amounts are entered by computer operator from image displayed on computer screen').

Given the 71 prose tasks on the combined DOL and NALS assessments, their type-of-information scores ranged from one (i.e., easiest) to five (i.e., most difficult). The mean type-of-information score was 3.00 (with a standard deviation of 1.21) and a median score of three. The mean values for type of requested information by Level were 1.83 for Level 1; 2.53 for Level 2; 2.88 for Level 3; 3.50 for Level 4; and 4.25 for Level 5.

The variable, *type of match*, refers to the processes used to relate information in a question or directive to corresponding information in the a text, and to the processes used to select a response from a range of response options (Kirsch et al., 1994). Type of match is influenced by the following conditions (see Figure 2 below for a summary). On average, "locate" matches are easier than "cycle" matches, cycle matches are easier than "integrate" matches, and integrate matches are easier than "generate" matches. In *locate tasks*, users match one or more features in a question or directive to one or more

features in the text (Kirsch & Mosenthal, 1992). Based on this match, the answer is located in the "node" (i.e., a paragraph or sentence) associated with these features.

In *cycle tasks*, users perform an iterative series of locate matches (Mosenthal & Kirsch, 1992). Cycle tasks may involve the selection of information that meets a particular criterion or condition (e.g., identify three different sources of lead poisoning throughout an article). Cycle tasks are further made difficult depending upon whether they are performed *within a paragraph* or *between paragraphs*.

Integrate tasks require users to compare or contrast information that has been located in two or more different locate matches or in one or more cycle matches (Kirsch & Mosenthal, 1992/1993). In general, integrate tasks which require readers to compare information are easier than those require readers to contrast information.

In *generate tasks*, readers are required to use prior knowledge (often representing a specific type of content knowledge) to match information in a question or directive to corresponding information in text (Mosenthal & Kirsch, 1993a). Moreover, generate tasks may require readers to use specialized knowledge to select from among a set of plausibly correct responses the answer which best meets the conditions stated in a question or directive.

In prose, matching is further made difficult as the *number of phrases* which make up a question or directive increases. In this case, "number of phrases" refers to the number of dependent and independent clauses which comprise a question or directive. For instance, a question consisting of only one independent clause is, on average, easier to answer than a question consisting of one independent clause and one dependent clause. In turn, a question consisting of one independent clause and two dependent clauses is, on average, even harder.

Matching is further made difficult depending upon the *number of responses users must supply* and *whether or not the number of responses, if greater than one, is specified in the question or directive*. Questions and directives requiring readers to list

only one answer are easier than those requiring readers to list two or three answers; questions and directives requiring readers to list two or three answers are easier than those requiring readers to list four answers. Questions and directives which specify the number of multiple responses to be listed are easier than those which do not specify the number of responses explicitly.

In addition, matching is made difficult to the extent that users have to make *inferences* either to match information in the question or directive to information in text. In such cases, low text-based inferences (or inferences which can be made within the context of information provided in a text) are easier to make than high text-based inferences (or inferences which require some combination of knowledge of the text and specialized prior knowledge).

Similarly, type of match is made difficult to the extent that, once a match has been made, readers then have to choose between two or more pieces of information in order to complete the new (or requested) information frame and, thus, answer the question. In some instances, readers may have to make a low or high text-based inference to determine why one of several possible answers best completes a requested information frame; or readers may have to identify conditional information which renders one answer more consistent with the conditions of choice established by a question or directive. Still in other cases, readers may have to relate a pronoun to its antecedent before an answer can be provided.

Based on the preceding observations, the following list of rules, shown in Figure 2, were used to score for type of match in this study. Note that these rules are specified additively. In other words, a task might be assigned a difficulty score of "7" because: It is a low text-based (add 1) cycle (add 2) task between paragraphs (add 1) which involves a two clause question (add another 1) that requires readers to list two responses (add another 1) whose actual number is explicitly specified in the question or directive (add another 1) (type-of-match score total = 7).

Insert Figure 2 about here.

Also note that, while the scoring system for type of match (as presented in Figure 2) could theoretically generate scores which could go as high as 20, this was not the case in the combined DOL and NALS assessments. Rather difficulty scores for type of match ranged from one to seven. The ceiling of seven was not set arbitrarily. Rather, this ceiling reflects the range of difficulty combinations which commonly characterize tasks found in society and the workplace (Paincaud & Jezak, 1994). While more difficult tasks could be conceived in designing assessments (e.g., a four phrase contrast task requiring high text-based inferencing and six uncued responses), such tasks would indeed be extremely artificial and would bear little resemblance to those tasks associated with every-day prose use (Kirsch & Jungeblut, 1986; Kirsch et al., 1992, 1993).

In sum, given the 71 prose tasks on the combined DOL and NALS assessments, their type-of-match scores ranged from one (i.e., easiest) to seven (i.e., most difficult). The mean type-of-match score was 3.66 (with a standard deviation of 1.49) with a median score of four. The mean values for type of match by Level were 1.00 for Level 1; 2.71 for Level 2; 3.58 for Level 3; 4.94 for Level 4; and 5.38 for Level 5.

In addition to type of information and type of match, a third process variable is *plausibility of distractors* (Kirsch et al., 1994). This variable has to do whether or not there are features from a question or directive's given and/or requested information which appear in the text but, once matched or identified, do not yield the correct requested information. Based on previous research, Kirsch et al. (1994) found that tasks are easiest to process when there are no plausible distractors in a text. (In this study, such tasks weree assigned a score of 1 for plausibility of distractors.) This is often the case when there is no other information related to the conditions set forth by a question or directive other than the answer.

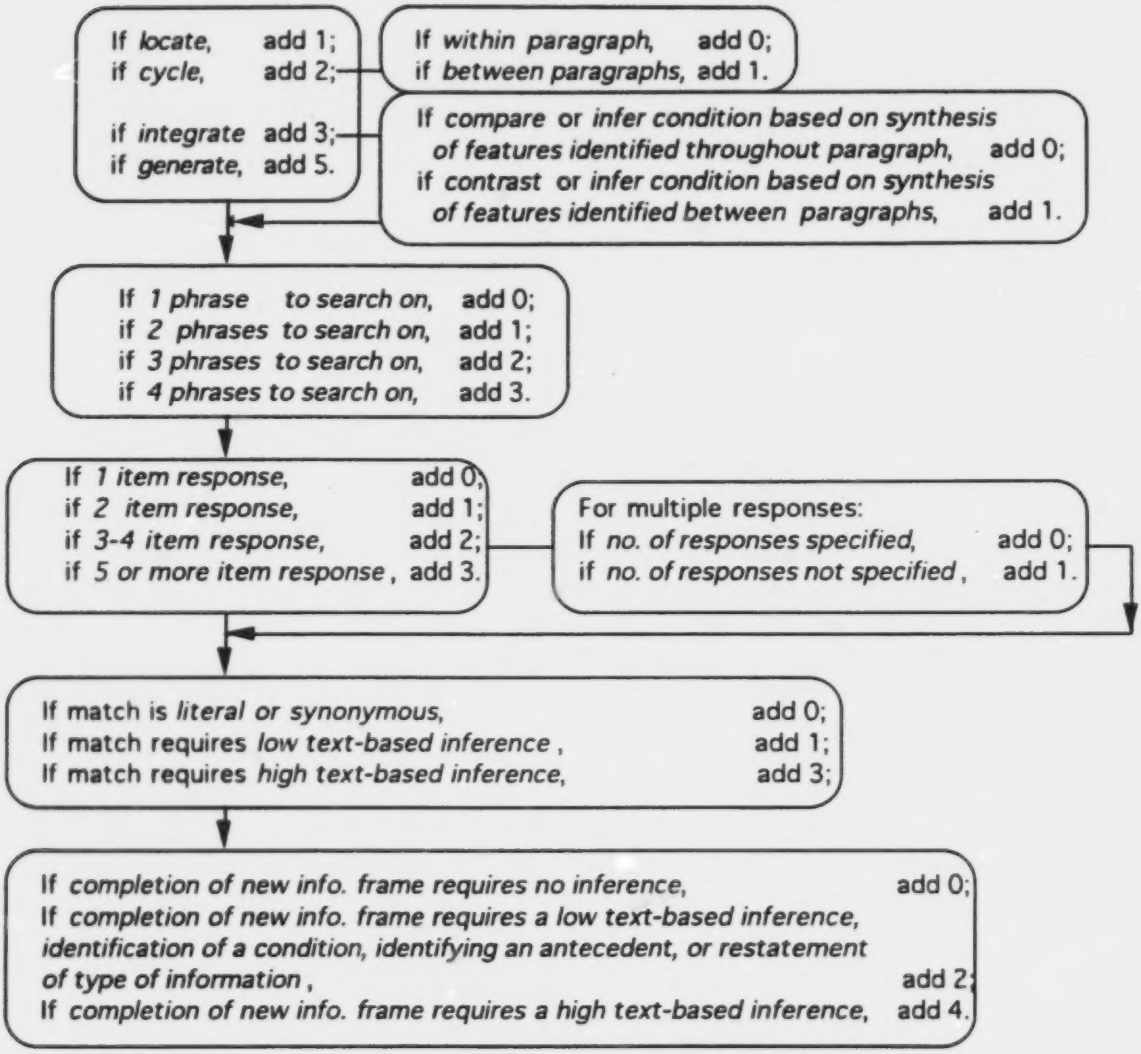

If locate, add 1;
if cycle, add 2;
if integrate add 3;
if generate, add 5.
If within paragraph, add 0;
if between paragraphs, add 1.
If compare or infer condition based on synthesis of features identified throughout paragraph, add 0;
if contrast or infer condition based on synthesis of features identified between paragraphs, add 1.
If 1 phrase to search on, add 0;
if 2 phrases to search on, add 1;
if 3 phrases to search on, add 2;
if 4 phrases to search on, add 3.
If 1 item response, add 0;
if 2 item response, add 1;
if 3-4 item response, add 2;
if 5 or more item response, add 3.
For multiple responses:
If no. of responses specified, add 0;
if no. of responses not specified, add 1.
If match is literal or synonymous, add 0;
If match requires low text-based inference, add 1;
If match requires high text-based inference, add 3;
If completion of new info. frame requires no inference, add 0;
If completion of new info. frame requires a low text-based inference, identification of a condition, identifying an antecedent, or restatement of type of information, add 2;
If completion of new info. frame requires a high text-based inference, add 4.

Figure 2. Additive Scoring Rules for Type of Match in Prose Processing

Tasks become slightly more difficult when: (a) plausible distractors for either given or requested (but not both) appear in a paragraph (including the paragraph in which the answer is located), or (b) when plausible distractors arise as invited inferences based on information in the paragraph in which the answer occurs. (In this study, such tasks were assigned a score of **2** for plausibility of distractors.)

Tasks become more difficult when plausible distractors for both given and requested information appear in different nodes, one of which may be in the paragraph in which the answer occurs. (In such cases, tasks in this study were assigned a score of **3** for plausibility of distractors.)

Tasks continue to increase in difficulty when: (a) plausible distractors for given and requested information both appear in the same paragraph but other than the paragraph in which the answer appears, or (b) when plausible distractors represent the opposite condition of what is established in the question or directive, and these distractors appear in a paragraph other than the one in which the answer occurs. (In such cases, tasks were assigned a score of **4** for plausibility of distractors.)

Finally, tasks are most difficult when: (a) plausible distractors for given and requested information both appear in the same paragraph as the answer, or (b) when plausible distractors represent the opposite condition of what is established in the question or directive, and these distractors appear in the same paragraph in which the answer occurs. (In this study, such tasks received a score of **5** for plausibility of distractors.)

Given the 71 prose tasks on the combined DOL and NALS assessments, plausibility-of-distractors scores ranged from one (i.e., easiest) to five (i.e., most difficult). The mean plausibility-of-distractors score was 2.59 (with a standard deviation of 1.26) with a median score of two. The mean values for plausibility of distractors were 1.17 for Level 1; 1.94 for Level 2; 2.46 for Level 3; 3.44 for Level 4; and 3.75 for Level 5.

To ensure reliability, two raters independently scored all the tasks comprising the DOL and NALS prose scales in terms of readability, type of information, type of match, and plausibility of distractors. There was 95 percent agreement on readability, 90 percent agreement on type of information, 83 percent agreement on type of match, and 86 percent on plausibility of distractors. Differences between raters were discussed and were agreed upon through consensus.

Defining and Illustrating the Five Levels of Prose Proficiency

After each of the variables had been scored in terms of their difficulty (as shown in Table 2), we then looked for patterns of similarity among their construct characteristics. (In anthropometry, this would be the same as determining the best combinations of sleeve, neck, and waist sizes so that the categories "small," "medium," "large," "extra large," and "extra-extra large" would account for the greatest variance in human size and shape as possible.) We describe and illustrate these patterns below, discussing the variables and constructs associated with each Level.

Prose Level 1. As shown in Table 2, we identified tasks in Level 1 as those which range below 225 in RP80 value. The process variables in this range tend to include combination scores of 1, 1, 2 or less. This combination accounted for five out of the six tasks within this Level (or 83 percent). These tasks had an average RP80 value of 197. Approximately twenty-one percent of the adults in the U. S. perform at this Level (Kirsch et al., 1993).

Most of the Tasks at this Level required readers to identify information which is quite concrete, including a 'person,' 'place,' or 'thing,' as well as an "attribute,' 'amount,' 'type of,' 'temporal,' 'action,' 'procedure,' or 'location.' Moreover, to complete these tasks, readers must process relatively short text to locate a single piece of information which is identical to (or synonymous with) the information given in the question or directive. If distractors appear in the text, they tend be located in a paragraph other than the one in which the correct answer occurs.

An example of a Level 1 task was based on stimulus A below. This text consists of 160 words and reflects an eighth-grade readability level. The directive related to this text instructed readers to "Underline the sentence that tells what Ms. Chanin ate during the swim." To complete this directive, readers have to recognize that the requested information is a "thing" (i.e., food). To identify this information, readers must make a synonymous match between 'ate' in the directive and "'banana and honey' sandwiches, hot chocolate, lots of water, and granola bars" in the text. Note that, since there is no other mention of food in the text, there are no plausible distractors for requested information.

Insert A about here.

Prose Level 2. We identified tasks in Level 2 as those which range between 226 and 275 in RP80 value (see Table 2). The process variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 1. Thus in Level 2, we find process combination scores of: 2, 2, 2; or 3, 3 or less, 3 or less. These combinations accounted for 12 out of 17 tasks within this level (or 71 percent). These tasks had an average RP80 value of 253. Twenty-seven percent of the adults in the U. S. perform at this Level with 80 percent probability (Kirsch et al., 1993).

Like tasks in Level 1, many tasks in Level 2 ask readers to complete information which is fairly concrete. However, in Level 2, we find some tasks which also require readers to identify information representing 'manner,' 'goal,' 'purpose,' 'attempt,' 'alternative,' and 'condition' information. Moreover, tasks at Level 2 often require readers to make a low level inference, or recognize a condition or an antecedent in order to identify requested information in a text. Finally, tasks at this Level tend to have a distractor for either given or new information present but not in the same paragraph as the answer.

A

Swimmer completes Manhattan marathon

The Associated Press

NEW YORK—University of Maryland senior Stacy Chanin on Wednesday became the first person to swim three 28-mile laps around Manhattan

Chanin, 23, of Virginia, climbed out of the East River at 96th Street at 9:30 p m She began the swim at noon on Tuesday

A spokesman for the swimmer, Roy Brunett, said Chanin had kept up her strength with "banana and honey" sandwiches, hot chocolate, lots of water and granola bars "

Chanin has twice circled Manhattan before and trained for the new feat by swimming about 28 4 miles a week The Yonkers native has competed as a swimmer since she was 15 and hoped to persuade Olympic authorities to add a long-distance swimming event

The Leukemia Society of America solicited pledges for each mile she swam

In July 1983, Julie Ridge became the first person to swim around Manhattan twice With her three laps, Chanin came up just short of Diana Nyad's distance record, set on a Florida-to-Cuba swim

Reduced from original copy

BEST COPY AVAILABLE

A task representative of this Level had a difficulty value of 250 and, again, was based on the stimulus **A** (shown above). This task included the question, "At what age did Chanin begin swimming competitively?" To answer this question, readers must recognize that the requested information is an amount (i.e., age). To identify this information, readers have to make a synonymous match between 'competitively' in the question and 'compete' in the text. Having made this match, readers then must find the answer '15' in the sentence, "The Yonkers native has competed as a swimmer since she was 15 and hoped to persuade Olympic authorities to add a long-distance swimming event." What makes this task somewhat difficult is the fact that there is a distractor for requested information which appears earlier in the text; this distractor includes the information that the swimmer's current age is 23.

Another task found at Level 2 on both the DOL and NALS prose scales had a difficulty value of 275. This task is shown in **B** below. This task instructed readers to "Circle the letter next to the note which best follows the instructions supplied by the company." To complete this task, readers have to identify a particular condition (i.e., what is specifically wrong with the appliance) as type of requested information. To carry out this identification, readers must make a low level inference that this condition is best met by the description "(My clock radio) rings 15-30 minutes later." In addition, note that this task involves the process of matching the pronoun 'it' with its antecedent 'clock radio.' This task is further complicated by the fact that there are other distractors in **B** which also allude to what is wrong with the clock radio but in a very general manner (e.g., "The clock does not run correctly on this clock radio.")

Insert **B** about here.

Prose Level 3. We identified tasks in Level 3 as those which range between 276 and 325 in RP80 value. The process variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 2.

B

A manufacturing company provides its customers with the following instructions for returning appliances for service:

> When returning appliance for servicing, include a note telling as clearly and as specifically as possible what is wrong with the appliance

A repair person for the company receives four appliances with the following notes attached. Circle the letter next to the note which best follows the instructions supplied by the company.

A
> The clock does not run correctly on this clock radio. I tried fixing it, but I couldn't.

B
> My clock radio is not working. It stopped working right after I used it for five days.

C
> The alarm on my clock radio doesn't go off at the time I set. It rings 15-30 minutes later.

D
> This radio is broken. Please repair and return by United Parcel Service to the address on my slip.

BEST COPY AVAILABLE

Thus in Level 3, we find process combination scores with one variable having a value of 4, 3 or less, 3 or less. This combination accounted for 19 out of 24 tasks within this level (or 79 percent). These tasks had an average RP80 value of 296. Thirty-two percent of the adults in the U. S. perform at this Level with 80 percent probability (Kirsch et al., 1993).

Tasks at this Level tend to require readers to identify condition information. In other instances, these tasks require readers to identify a 'reason' or 'explanation.' In terms of type of match, Level 3 tasks again require readers to make literal, synonymous, and low level inference matches between the question or directive and the text. However, unlike Level 1 and 2 locate tasks, Level 3 tasks usually require readers to identify and list multiple responses (the number of which is specified in the question or directive). Also the questions and directives of Level 3 tasks tend to consist of several phrases. Moreover, these tasks generally require readers to complete requested information by identifying special conditional information stated in a question or directive or by establishing antecedence between a pronoun and its reference. Distracting information for both given and requested information tends to be present, both of which appear in different paragraphs from one another and neither of which appear in the same paragraph as the answer.

One of the more difficult Level 3 tasks (with an RP80 value of 316) involved the stimulus shown in **C**. This text consists of 688 words and has a seventh-grade readability level. The directive associated with this stimulus asked readers to "List two things that Chen became involved in or has done to help resolve conflicts due to discrimination." Note that this directive consists of three phrases: 'Chen became involved in,' 'has done,' and 'to resolve conflicts due to discrimination.' To carry out this directive, readers must recognize that the requested information is a set of actions.

To identify these actions, readers must match 'resolve conflicts' in the directive with 'resolving . . . conflicts' in the third to last paragraph. Having made this locate

match, readers must then make two low text-based inferences (as well as establish an antecedent reference) to identify Chen's actions in response to discrimination. These actions are stated in the paragraph in which the original match was made and in the next contiguous paragraph. Having made these inferences, readers would list "Chen called for a meeting with the leaders of the Korean community to help resolve the conflict" and "Chen has been involved in Hispanic, Jewish and Black issues . . ." as the correct answers.

Insert **C** about here.

Prose Level 4. We identified tasks in Level 4 as those which ranged between 326 and 375 in RP80 value. The process variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 3. Thus in Level 4, we find process combination scores 4, 4 or less, 4 or greater. This combination accounted for 11 out of 16 tasks within this Level (or 69 percent). These tasks had an average RP80 value of 350. Seventeen percent of the adults in the U. S. perform at this Level with 80 percent probability (Kirsch et al., 1993).

Tasks at this Level tend to require readers to identify rather abstract information, including 'reason,' 'evidence,' 'explanation,' 'causation,' 'result,' 'comparison,' and 'contrast.' In terms of type of match, Level 4 tasks generally require readers to cycle and integrate, as well as locate. Again, multiple responses may be required but for which the number of responses is not specified. As with Level 3 tasks, Level 4 tasks often require readers to complete requested information by identifying special conditional information stated in a question or directive, or by establishing antecedence between a pronoun and its reference. In other cases, high text-based inferences must be made to distinguish the correct requested information from distracting information. At this Level, distracting information for both given and

C

IDA CHEN is the first Asian-American woman to become a judge of the Commonwealth of Pennsylvania.

She understands discrimination because she has experienced it herself.

Soft-spoken and eminently dignified, Judge Ida Chen prefers hearing about a new acquaintance rather than talking about herself. She wants to know about career plans, hopes, dreams, fears. She gives unsolicited advice as well as encouragement. She instills confidence.

Her father once hoped that she would become a professor. And she would have also made an outstanding social worker or guidance counselor. The truth is that Chen wears the caps of all these professions as a Family Court judge of the Court of Common Pleas of Philadelphia County, as a participant in public advocacy for minorities, and as a particularly sensitive, caring person.

She understands discrimination because she has experienced it herself. As an elementary school student, Chen tried to join the local Brownie troop. "You can't be a member," she was told. "Only American girls are in the Brownies."

Originally intent upon a career as a journalist, she selected Temple University because of its outstanding journalism department and affordable tuition. Independence being a personal need, she paid for her tuition by working for Temple's Department of Criminal Justice. There she had her first encounter with the legal world and it turned her career plans in a new direction — law school.

Through meticulous planning, Chen was able to earn her undergraduate degree in two and a half years and she continued to work three jobs. But when she began her first semester as a Temple law student in the fall of 1973, she was barely able to stay awake. Her teacher Lynne Abraham, now a Common Pleas Court judge herself, couldn't help but notice Chen yawning in the back of the class, and when she determined that this student was not a party animal but a workhorse, she arranged a teaching assistant's job for Chen on campus.

After graduating from Temple Law School in 1976, Chen worked for the U.S. Equal Employment Opportunity Commission where she was a litigator on behalf of plaintiffs who experienced discrimination in the workplace, and then moved on to become the first Asian-American to serve on the Philadelphia Commission on Human Relations.

Appointed by Mayor Wilson Goode, Chen worked with community leaders to resolve racial and ethnic tensions and also made time to contribute free legal counsel to a variety of activist groups.

The "Help Wanted" section of the newspaper contained an entry that aroused Chen's curiosity — an ad for a judge's position. Her application resulted in her selection by a state judicial committee to fill a seat in the state court. And in July of 1988, she officially became a judge of the Court of Common Pleas. Running as both a Republican and Democratic candidate, her position was secured when she won her seat on the bench at last November's election.

At Family Court, Chen presides over criminal and civil cases which include adult sex crimes, domestic violence, juvenile delinquency, custody, divorce and support. Not a pretty picture.

Chen recalls her first day as judge, hearing a juvenile dependency case — "It was a horrifying experience. I broke down because the cases were so depressing," she remembers.

Outside of the courtroom, Chen has made a name for herself in resolving interracial conflicts, while glorying in her Chinese-American identity. In a 1986 incident involving the desecration of Korean street signs in a Philadelphia neighborhood, Chen called for a meeting with the leaders of that community to help resolve the conflict.

Chen's interest in community advocacy is not limited to Asian communities. She has been involved in Hispanic, Jewish and Black issues, and because of her participation in the Ethnic Affairs Committee of the Anti-Defamation League of B'nai B'rith, Chen was one of 10 women nationwide selected to take part in a mission to Israel.

With her recently won mandate to judicate in the affairs of Pennsylvania's citizens, Chen has pledged to work tirelessly to defend the rights of its people and contribute to the improvement of human welfare. She would have made a fabulous Brownie.

— *Jessica Schultz*

BEST COPY AVAILABLE

requested information tends to be present, both of which may appear in the same paragraph as the answer.

An example of a Level 4 task (with an RP80 of 362) was one based on the stimulus shown in D. This stimulus consists of 362 words and represents a sixth-grade readability level. A question associated with this stimulus was "If you are working, you may be able to get supplemental security income as an individual if you earn less than what amount per month?" Together this question and its related stimulus comprised a task which had a difficulty value of 362. This directive consists of four phrases: 'If you are working,' 'you may be able,' 'to get supplemental security income as an individual,' 'if you earn less than what amount per months.'

To answer this question, readers must recognize that the requested information is an amount. To identify this amount, readers must make a series of cycle matches between paragraphs. In short, they first must match 'income' in the question to the heading 'income' in the text. Next, they must match 'If you are working' in the question to 'If you work' in the text, followed by the match 'individual . . . per month' in the question to 'a month for an individual' in the text. At this point, readers would identify the answer '$821.' In selecting this amount, readers would have to be careful not to inadvertently identify the plausible distractor '$1,191' which appears in the same paragraph as the correct answer and shares the same feature 'If you work' with this answer.

Insert D about here.

Prose Level 5. We identified tasks in Level 5 as those which are above 375 in RP80 value. The process variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 4. Thus in Level 5, we find process combination scores of 5, less than 5, greater than 5. This combination accounted for seven out of eight tasks within this Level (or 88 percent). These tasks had

ARE YOU ELIGIBLE FOR SSI?

WHAT IS SSI?

SSI stands for supplemental security income. It is a Federal program run by the Social Security Administration. It pays monthly checks to aged, blind, and disabled people who do not have much income or resources.

Under SSI, *aged* means you are 65 or older. *Blind* means the vision in your better eye is 20/200 or less or you have a limited visual field of 20 degrees or less. *Disabled* means you have a severe physical or mental condition that keeps you from doing any substantial gainful work, and medical evidence shows it is expected to last at least 12 months or result in death.

HOW MUCH IS AN SSI MONTHLY PAYMENT?

The basic Federal monthly payment is:

- $368 for an eligible person
- $553 for an eligible couple

But, you may not get this exact amount. You may get *less* if you, your spouse, or your parents (if you are under 18) have other income. Or you may get *more* if you live in a State that adds money to the Federal payment.

RESOURCES AND INCOME

To get SSI, your resources and income must be below certain amounts. Both resources and income are explained below.

RESOURCES

Resources are the things you own, such as real estate, personal belongings, cash, bank accounts, or stocks and bonds. We do not count all of your resources. For example, we do not count your home or some of your personal belongings. And we usually do not count your car.

You may be able to get SSI if the resources we count are less than:

- $2,000 for an individual
- $3,000 for a couple

INCOME

Income means earnings, Social Security checks, and pensions *plus* non-cash items you receive such as food, clothing, or shelter.

If you do not work, you may be able to get SSI if all your income adds up to less than:

- $388 a month for an individual
- $573 a month for a couple

If you work, you may be able to get SSI if your earnings are less than:

- $821 a month for an individual
- $1,191 a month for a couple

(These figures may be higher if you live in a State that adds money to the Federal payment.)

OTHER RULES YOU SHOULD KNOW

Before you can get SSI checks, you must meet these other requirements:

- Be a U.S. citizen, or a lawfully admitted immigrant, or an alien permanently residing in the U.S. "under color of law";
- Be a resident of the U.S. or the Northern Mariana Islands, and
- Apply for any other money benefits that may be due you.

ALSO: If you are disabled, you must accept vocational rehabilitation if offered to you. And if you are a disabled drug addict or alcoholic, you must accept available treatment and receive SSI payments through a representative payee.

OTHER BENEFITS

Being eligible for SSI also means you may be eligible for other benefits such as food stamps, Medicaid, and social services. For information about these benefits, ask the people at your local Social Security office.

HOW TO APPLY FOR BENEFITS

It is easy. Just call any Social Security office. You can find the phone number in the telephone directory under "Social Security Administration" or "U.S. Government."

BEST COPY AVAILABLE

an average RP80 value of 419. Three percent of the adults in the U. S. perform at this Level with 80 percent probability (Kirsch et al., 1993).

Tasks at this Level tend to require readers to identify quite abstract information, including 'contrast,' 'equivalence,' and 'theme' (or 'summary'). In terms of type of match, Level 4 tasks often require readers generate as well as locate, cycle, and integrate. Generate may involve the use of specialized background knowledge to interpret a phrase or to synthesize text information. Also at this Level, distracting information for both given and requested information may be present, both of which frequently appear in the same paragraph as the answer.

An example of a Level 5 task was based on the stimulus shown in E. This stimulus consists of 653 words and represents a seventh-grade readability level. A directive associated with this stimulus was "Identify and summarize the two kinds of challenges that attorneys use while selecting members of a jury." Together this directive and its related stimulus comprised a task which had a difficulty value of 410. This directive consists of two phrases: 'the two kinds of challenges that attorneys use' and 'while selecting members of a jury.'

To complete this directive, readers must recognize that the requested information involves contrasting two kinds of challenges. To accomplish this, readers must cycle, first matching 'challenges' in the question to 'challenge' in the sentences:

> When an attorney believes that there is a legal reason to excuse a juror, he or she will challenge the juror for cause. Unless both attorneys agree that the juror should be excused, the judge must either sustain or override the challenge.

Readers must next match 'challenges' in the question to 'challenge' in the sentences:

> After all challenges for cause have been ruled upon, the attorneys will select the trial jury from those who remain by exercising peremptory challenges. Unlike challenges for cause, no reason need be given for excusing a juror by peremptory challenge. Attorneys usually exercise these challenges by taking turns striking names from a list until both are satisfied with the jurors at the top of the list or until they use up the number of challenges allowed. Challenged jurors and any extra jurors will then be excused and asked to return to the jury selection room.

Having cycled and identified the two types of challenges, readers next must integrate the information by identifying (via inferencing) the points of contrast, as well as describing the critical features of these points of contrast (cf., Kirsch & Mosenthal, 1991). These points of contrast and their descriptive features would include such things as the 'presence or absence of a reason for issuing a challenge' (i.e., challenges for cause involve reason while peremptory challenges do not), 'the sequence of challenges' (i.e., challenges for cause precede peremptory challenges), and 'procedures for carrying out a challenge' (i.e., challenges for cause are arbitrated by a judge while peremptory challenges involve no such arbitration).

This task would have been easier had only one point of contrast and its descriptive features been required. If more points of contrast and their descriptive features had been required, the harder the task would have become, especially as no specific number contrasts are identified in the directive. Also, to the extent that jury selection is discussed in other paragraphs other than those in which the answer occurs, this information represents distracting given information.

Insert E about here.

Results

Correlations

In the above, four variables were described and illustrated which, based on previous research (Kirsch et al., 1994), have been shown to influence the difficulty of prose tasks. To examine the relations among these variables and task difficulty, we first computed intercorrelations between task RP80 and Level scores and the structure and process variables for the DOL and NALS prose task combined. These correlations are presented in Table 3 below.

E

DO YOU HAVE A QUESTION?

QUESTION: What is the new program for scheduling jurors?

ANSWER: This is a new way of organizing and scheduling jurors that is being introduced all over the country. The goals of this program are to save money, increase the number of citizens who are summoned to serve and decrease the inconvenience of serving.

The program means that instead of calling jurors for two weeks, jurors now serve only one day, or for the length of one trial if they are selected to hear a case. Jurors who are not selected to hear a case are excused at the end of the day, and their obligations to serve as jurors are fulfilled for three years. The average trial lasts two days once testimony begins.

An important part of what is called the One Day – One Trial program is the "standby" juror. This is a person called to the Courthouse if the number of cases to be tried requires more jurors than originally estimated. Once called to the Courthouse, the standby becomes a "regular" juror, and his or her service is complete at the end of one day or one trial, the same as everyone else.

Q. How was I summoned?

A. The basic source for names of eligible jurors is the Driver's License list which is supplemented by the voter registration list. Names are chosen from these combined lists by a computer in a completely random manner.

Once in the Courthouse, jurors are selected for a trial by this same computer and random selection process.

Q. How is the Jury for a particular trial selected?

A. When a group of prospective jurors is selected, more than the number needed for a trial are called. Once this group has been seated in the courtroom, either the Judge or the attorneys ask questions. This is called *voir dire*. The purpose of questions asked during *voir dire* is to ensure that all of the jurors who are selected to hear the case will be unbiased, objective and attentive.

In most cases, prospective jurors will be asked to raise their hands when a particular question applies to them. Examples of questions often asked are: Do you know the Plaintiff, Defendant or the attorneys in this case? Have you been involved in a case similar to this one yourself? Where the answer is yes, the jurors raising hands may be asked additional questions, as the purpose is to guarantee a fair trial for all parties. When an attorney believes that there is a legal reason to excuse a juror, he or she will challenge the juror for cause. Unless both attorneys agree that the juror should be excused, the Judge must either sustain or override the challenge.

After all challenges for cause have been ruled upon, the attorneys will select the trial jury from those who remain by exercising peremptory challenges. Unlike challenges for cause, no reason need be given for excusing a juror by peremptory challenge. Attorneys usually exercise these challenges by taking turns striking names from a list until both are satisfied with the jurors at the top of the list or until they use up the number of challenges allowed. Challenged jurors and any extra jurors will then be excused and asked to return to the jury selection room.

Jurors should not feel rejected or insulted if they are excused for cause by the Court or peremptorily challenged by one of the attorneys. The *voir dire* process and challenging of jurors is simply our judicial system's way of guaranteeing both parties to a lawsuit a fair trial.

Q. Am I guaranteed to serve on a jury?

A. Not all jurors who are summoned actually hear a case. Sometimes all the Judges are still working on trials from the previous day, and no new jurors are chosen. Normally, however, some new cases begin every day. Sometimes jurors are challenged and not selected.

BEST COPY AVAILABLE

Overall, comparing the relations between structure and process variables and task difficulty, the following was found. Type of match correlated highest with RP80 (.84) and Level (.82), followed by plausibility of distractors (.60 with RP80 and .61 with Level) and type of information (.55 with RP80 and .53 with Level). Readability correlated moderately with RP80 (.32) and Level (.32). Among the process variables, there was relatively high intercorrelation between type of match and type of information (.54) but moderate correlation between type of match and plausibility of distractors (.38). The correlation between plausibility of distractors and type of information was only .03. Between the structure and process variables, correlations ranged from low (.13) between readability and type of information to moderate (.36) between readability and plausibility of distractors. The correlation between readability and type of match was .25.

Insert Table 3 about here.

Regression analyses

Next, two general multiple regression analyses were run using RP80 and Level as measures of task difficulty. Table 4 shows the results of these analyses. The numbers in the table represent the raw beta coefficients for each of the variables included in the regression analyses. In addition, standard errors and *p* -values for each variable are listed. Overall, all three process variables were significant for both RP80 and Level ($p < .01$). In the full regression equation, readability proved nonsignificant ($p > .05$) for both RP80 and Level.

Insert Table 4 about here.

As shown in Table 4, the combined variables accounted for 83 percent of the *R*-squared variance when difficulty was defined using RP80 values, and 80 percent of the

Table 3

Intercorrelations between Prose Task Difficulty (Represented by RP80 and Level), and Structure and Process Variables for DOL and NALS Combined

	Task difficulty		Structure Variables	Process Variables	
	1	2	3	4	5
Task Difficulty					
1. RP80	-				
2. Level	.95				
Structure Variable		-			
3. Readability	.32	.32			
Process Variables			-		
4. Type of Match	.84	.82	.25	-	
5. Plausibility of Distractors	.60	.61	.36	.38	-
6. Type of Information	.55	.53	.13	.54	.03

Table 4

Raw Beta Coefficients and Standard Errors of Predictive Variables in Regression on Prose Task Difficulty Defined Using RP80 and Level for DOL and NALS Combined

	RP80			Level		
	Beta	Std. Error	p^1	Beta	Std. Error	p^1
Structure Variables						
Readability	.36	1.52	.80	.01	.03	.77
Process Variables						
Type of Match	23.96	2.77	.00	.41	.06	.00
Plausibility of Distractors	18.76	2.89	.00	.34	.06	.00
Type of Information	11.90	3.17	.00	.20	.06	.00
Total variance accounted for:						
R^2		84%			80%	
Adjusted R^2		83%			79%	

[1] df = 66

R-squared variance when difficulty was defined using Level. When entered into the regression equation by itself, readability was significant ($p < .02$), but accounted only for 10 percent of the *R*-squared variance for both RP80 and Level. However, entering this variable in the general regression equation with the three process variables did not increase explained *R*-squared variance for either RP80 or Level. As such, the results of these analyses attest to the importance of process variables over readability as being the better predictors of prose task difficulty on the combined DOL and NALS assessments.

Document Literacy

An additional aspect of being literate in today's society is having the knowledge and skills needed to process documents, or information organized in matrix structures (i.e., in rows and columns). Included among documents are such things as tables, indexes, lists, coupons, schedules, charts, graphs, maps, and forms. In contrast to prose, which tends to be the predominant form of literacy in schools, documents tend to be the principal form of literacy in non-school settings (Guthrie, Seifert, & Kirsch, 1986). Documents serve many important functions in our daily lives (Bassett, Goodman, & Fosegan, 1981; Burch & Grudnitski, 1986). They enable people to perform important actions (e.g., applying for benefits, opening a charge account), make informed decisions (e.g., using a table of benefits to determine whether certain medical costs are covered), and record actions (completing a deposit slip or bill of sale, receiving a ticket for speeding).

Besides being important to our daily lives, documents are extremely pervasive in our information-rich society (Hartley, 1985; Rayner, 1982). Rayner (1982), for example, estimated the total number of different British government forms to be well over 100,000. This figure included about 38,000 external forms—those issued to the public or to businesses—and about twice as many internal administrative forms. The Associated Press (Miller, 1984) estimated that, in the mid-1970s, the United States government issued over 98,000 different forms per year and received over 500 million

responses. During this period, the Internal Revenue Service alone sent out over 3,500 different forms. Given the increase in information necessary to maintain the various organizations of our complex society, the number of documents issued will most likely continue to increase dramatically over the years to come (Burch & Grudnitski, 1986).

Document Variables

Like prose tasks, document tasks require readers to identify requested information in terms of different degrees of abstractness. This is represented by the variable "type of information." Moreover, both require readers to match information in a question or directive to corresponding information in a text or document. This again involves the strategies of locating, cycling, integrating, and generating information; these processes (as in prose) are represented by the variable "type of match." Finally, prose and document tasks both may be made more difficult due to the presence of distracting information. This is represented by the variable called "plausibility of distractors."

While these variables are similar in name to those which influence prose task processing, they are different in document task processing since they are implemented in the context of information organized in matrix rather than paragraph structures. In some instances, information may be organized as a *simple list* comprised of a set of items and a label (e.g., a list of grocery items labeled 'food to buy') (Mosenthal & Kirsch, 1989b). Moreover, this information may be organized as a *combined list* in which one column of information often acts as a subject (e.g., 'U. S. Presidents') and additional columns concatenated to the subject column function as predicates (e.g., 'places of birth,' 'date of inauguration,' and 'date of death') (Kirsch & Mosenthal, 1989). In other instances, this information may be organized as an *intersected list* (e.g., a TV schedule) where the items in one list (i.e., the intersected list such as shows) are concatenated with a row list (e.g., the intersecting list of times) and a column list (e.g., the intersecting list of channels) (Mosenthal & Kirsch, 1989c). Finally, document

information may be organized as a *nested list* in which two or more lists with the same labels are embedded under different lists (e.g., two lists detailing the amount of sales by sales person are nested by sales quarter) (Kirsch & Mosenthal, 1990b).

These four variables are described below. Using these variables, we identify construct patterns characteristic of document task difficulty within each of the five Levels (see Table 5 for a listing of the document tasks and their variable scores). We illustrate these Levels with representative tasks . Finally, we close this section by presenting the statistical significance of our variables as predictors of task difficulty..

Structural complexity. The basic structural unit of documents are "simple lists" (Mosenthal & Kirsch, 1989b). As noted above, such lists consist of a series of exemplars, or items, which belong to a common class of elements (e.g., kinds of materials, types of operations, various conditions). In many instances, these items are organized in terms of a more generic category called a "label."

In analyzing the structural complexity of documents used in the DOL and NALS assessment, we first divided the document stimuli into their respective simple lists. (In the case of those documents where simple lists consisted of a series of sentences, items were defined simply as the number of dependent and independent clauses which comprised these sentences.) Next, the number of items in each list were totaled and recorded, as was the number of explicit labels. As shown in Table 5, the stimuli comprising the NALS document scale ranged from zero (consisting of all labels, as is characteristic of some forms) to 758 items, and from zero to 180 labels.

Insert Table 5 about here.

Based on the distribution of items and labels as well as document types across the full range of tasks on the document scale, we created the following document "readability" variable. This variable combines type of document with the number of items and labels comprising a document (Mosenthal & Kirsch, in press). For type of

Table 5

Characteristics of Variables (i.e., Readability ('Read'), Type of Match ('TOM'), Plausibility of Distractors ('POD') and Type of Information ('TOI') for Document Tasks by RP80 and Difficulty Level for DOL and NALS Combined

Description	RP80	Level	Read	TOM	POD	TOI
Social Security Card*	69	1	1	1	1	1
Driver's License*	178	1	2	1	2	1
Traffic Signs	178	1	1	1	1	1
Room Preparation Form*	180	1	1	1	1	1
Room Preparation Form*	187	1	1	1	1	1
Medicine Dosage	187	1	2	1	2	2
TV Schedule*	188	1	8	2	2	1
Registration & Tuition Info	190	1	3	1	2	2
Theater Trip Information	198	1	2	1	2	2
Room Preparation Form*	198	1	1	1	2	1
Phone Message*	199	1	1	1	2	1
Phone Message*	202	1	1	1	1	1
How Companies Share Market	203	1	7	2	2	2
Food Coupons	205	1	3	2	2	1
Room Preparation Form*	205	1	1	2	2	1
Essence Table of Contents	211	1	3	1	2	2
MasterCard/Visa Statement	212	1	6	1	2	2
Black/White Middle-Class	213	1	4	3	2	2
Recreation Vehicles	214	1	2	1	2	2
Room Preparation Form*	216	1	1	1	2	2
Dessert Recipes	217	1	5	3	2	1
Deposit Slip*	223	1	3	2	2	1
Deposit Slip*	223	1	3	1	1	1
Wage and Tax Statement*	224	1	5	2	2	2
El Paso Gas & Electric Bill	224	1	8	1	2	2
Classified Ads*	228	2	8	2	3	1
Mercer County Map*	232	2	4	2	2	2
Bus Schedule	233	2	2	2	3	2
Public School Handbook Info*	233	2	6	2	2	3
Maintenance Record	233	2	3	2	3	2
Tempra Dosage Chart	234	2	5	2	3	3
Facts about Fire*	235	2	1	2	3	2
Phone Message*	237	2	1	2	3	1
Bennetts' Bill & Check*	238	2	6	3	2	1
Sign Out Sheet*	238	2	2	3	2	1
Community College Map	238	2	7	3	2	1
Dessert Recipes	239	2	5	3	2	1
Sign Out Sheet*	240	2	2	2	2	3
Social Security Application	242	2	5	2	3	1
Social Security Application	242	2	5	2	2	3
Certified Mail Label	243	2	2	2	2	2
How Companies Share Market	243	2	7	3	2	2
Bennetts' Bill & Check*	245	2	6	3	2	1
Bennetts' Bill & Check*	245	2	6	3	2	1
Employee Benefits Chart*	245	2	4	2	2	2
Bennetts' Bill & Check*	245	2	6	3	2	1
Catalog Page Order Form*	246	2	8	3	2	1
Phone Message*	249	2	1	2	2	2
Certified Mail Label	251	2	2	2	2	2
Social Security Application	251	2	5	2	2	1
College Football Awards*	254	2	6	2	2	3
Power Consumption Graph*	255	2	4	3	2	1
MasterCard/Visa Statement	256	2	6	1	2	2
El Paso Gas & Electric Bill	257	2	8	2	2	2

BEST COPY AVAILABLE

Essence Table of Contents	257	2	3	2	2	3
Social Security Application	259	2	5	2	2	2
Highland Route/Schedule	260	2	9	3	2	1
Consumer Reports Books	260	2	4	3	2	1
Pest Control Warning	260	2	2	3	3	2
Minimum Wage Power	260	2	4	3	3	2
Sales Graph by Season*	260	2	5	3	2	2
Yellow Pages: Microscope*	265	2	7	3	2	1
Trends	266	2	3	3	2	1
Community College Map	267	2	7	3	2	1
Deposit Slip*	271	2	3	3	3	2
Wage and Tax Statement*	275	2	5	2	3	2
Outland Award:*	275	2	5	3	2	2
El Paso Gas & Electric Bill	275	2	8	3	3	2
Registration & Tuition Info	276	3	3	2	3	2
Power Consumption Graph*	277	3	4	4	2	1
Abrasive Selection Guide*	280	3	10	3	2	3
Burning Out of Control	281	3	2	3	2	2
Washington/Boston Sched	283	3	9	2	4	2
City Hall Map*	284	3	4	4	2	2
Essence Table of Contents	285	3	3	4	2	3
Vitamin Page in Almanac*	287	3	5	3	4	2
Yellow Pages: Milk*	289	3	7	2	3	1
Hendricks Mining*	290	3	7	2	4	2
Phone Message*	292	3	1	3	2	1
Sign Out Sheet*	296	3	2	2	1	4
Abrasive Selection Guide*	303	3	10	4	2	2
Parents/Teachers Evaluate	304	3	4	3	2	4
Highland Route/Schedule	305	3	9	4	2	2
Power Consumption Graph*	307	3	4	4	2	1
Phone Message*	309	3	1	3	2	3
U. S. Savings Bonds	309	3	6	4	4	2
Consumer Reports Books	311	3	4	3	3	2
Almanac: Degree Celsius*	313	3	5	4	3	2
Vista Grande Bus Schedule*	313	3	10	3	3	2
Catalog Page Order Form*	313	3	8	4	4	3
Vista Grande Bus Schedule*	314	3	10	3	4	2
Trends	317	3	3	3	3	2
US Petroleum Imports*	318	3	10	6	2	2
MasterCard/Visa Statement	319	3	6	2	4	2
Maintenance Record	323	3	3	3	4	2
Vista Grande Bus Schedule*	324	3	10	3	4	2
Tempra Dosage Chart	327	4	5	4	4	2
Telephone Bill*	328	4	7	4	4	2
Consumer Report Index	331	4	7	3	5	3
Comparing Credit Cards	334	4	5	4	4	1
Parents/eachers Evaluate	342	4	4	2	4	2
Almanac: Petroleum Imports*	346	4	5	4	3	2
Vista Grande Bus Schedule*	348	4	10	3	4	2
MasterCard/Visa Statement	364	4	6	3	4	2
El Paso Gas & Electric Bill	364	4	8	5	4	5
Consumer Report Index	367	4	7	4	3	3
US Petroleum Imports*	378	5	11	7	2	3
US Exports/Imports*	379	5	7	3	3	3
Spolight Economy	383	5	10	3	3	2
Spolight Economy	386	5	10	3	3	3
Trends	386	5	3	3	3	3
Comparing Credit Cards	387	5	5	8	2	3
Parents/Teachers Evaluate	396	5	4	8	3	3
Spolight Economy	409	5	10	3	3	2
Income Tax Table	421	5	9	4	3	2
Spolight Economy	470	5	10	7	3	2

*Tasks on both the DOL and NALS assessments.

BEST COPY AVAILABLE

document, we scored **1** if it had a simple list structure; **2** if it had an underlying combined list structure; **3** if it had an underlying intersected list structure; **4** if it had an underlying nested list structure; and **5** if it was comprised of different multiple documents (e.g., three line graphs representing different measures of economic recovery) or if the document represented a poorly formed combined list with information concatenated in a "run-on" list structure (e.g., a paragraph list).

In terms of number of items, we scored **0** if the document consisted of 75 items or less; **1** if it consisted of 76 to125 items; **2** if it consisted of 126 to 175 items; and **3** if it consisted of 176 items or more. In terms of number of labels, we scored **0** if the document consisted of 15 labels or less; **1** if it consisted of 16 to 25 labels; **2** if it consisted of 26 to 35 labels; and **3** if it consisted of more than 35 labels.

After calculating each of these structural dimensions, we then added them to create a readability variable which ranged from one (i.e., the simplest document structure) to 11 (i.e., the most complex document structure). As shown in Table 5, the overall readability of documents on the combined DOL and NALS scales ranged from one to 11. The mean readability was 4.84 (with a standard deviation of 2.74) and a median of five. The mean readability by Level was as follows: 2.81 for Level 1; 4.58 for Level 2; 5.64 for Level 3; 6.30 for Level 4; and 7.50 for Level 5.

Process variables. The second set of variables which has been shown to influence document task difficulty include three process variables (Kirsch & Mosenthal, 1990a; Mosenthal & Kirsch, 1993b; Mosenthal & Kirsch, in press): "type of information," "type of match," and "plausibility of distractors." *Type of information* refers to the kind of information which users must identify to complete a question or directive. As Mosenthal and Kirsch (1991a) have noted, documents typically consist of a rather restricted range of information types. These information types form a continuum of concreteness which are operationalized as follows.

Questions requesting information regarding the identification of *persons, animals,* or *things* (e.g., "What requires the use of extra course production sandpaper to remove paint?" (answer: 'wood' and 'metal')) were scored the highest (i.e., **1**) in terms of concreteness. Questions requesting information regarding the identification of *amounts, times, attributes, types, actions,* and *locations* (e.g., "Medium production sandpaper is recommended for what type of stock removal?" (answer: 'moderate stock removal')) were assigned a concreteness score of **2**.

Questions requesting information regarding the identification of *manner, goal, purpose, alternative, attempt, condition, pronominal reference,* and *predicate adjectives* (e.g., "According to the safety information in the abrasive selection guide, when should one follow the manufacturer's recommended procedures?" (answer: 'when using power tools')) were assigned a concreteness score of **3**. Questions requesting information regarding the identification of *cause, effect, reason, result, evidence, similarity,* and *explanation* (e.g., "According to the safety information in the abrasive selection guide, what are two similarities between wood and metal in the use of production sandpaper?" (answer: 'Both require the use of extra coarse and coarse types of sandpaper to remove paint and stock')) were assigned a concreteness score of **4**. And finally, questions requesting information regarding the identification of *equivalent, difference, theme* (or *pattern*) were assigned a concreteness score of **5** ("equivalence," as in prose, consisted of a highly unfamiliar term for which readers had to provide a definition, e.g., "What type of material is 'heavy stock' as listed in the abrasive selection guide?").

Given the 117 document tasks on the combined DOL and NALS assessments, their type-of-information scores ranged from one (i.e., easiest) to five (i.e., most difficult). The mean type-of-information score was 1.99 (with a standard deviation of .98) with a median score of 2. The mean values for type of requested information were 1.46 for Level 1; 1.74 for Level 2; 2.14 for Level 3; 2.40 for Level 4; and 3.60 for Level 5.

The variable *type of match* refers to the processes required to relate information in the question or directive to corresponding information in the document and to the process of entering a response (Kirsch & Mosenthal, 1990a; Mosenthal & Kirsch, in press). Type of match is influenced by the following conditions (see Figure 3 below for a summary). On average, "locate" matches are easier than "cycle," cycle matches are easier than "integrate," and integrate matches are easier than "generate." In *locate tasks*, users match one or more features in a question or directive to one or more features in the document (Kirsch & Mosenthal, 1992). Based on this match, the answer is located in the "node" (i.e., either a matrix cell within a list, a matrix cell defined by the intersection of two or more lists, or a list itself) associated with these document features.

In *cycle tasks*, users perform an iterative series of locate matches, within a given list or between lists (Mosenthal & Kirsch, 1992). Cycle tasks within lists often involve the selection of items that meet a particular criterion (e.g., all states which have a minimum driving age of 18). Other cycle tasks require users to first locate information in one document and then, matching on the answer found in this list, make a new locate match (e.g., identifying which bar on a bar graph represents the lowest crime rate and then identifying what state this bar refers to in the chart legend). This may or may not result in the identification of the final answer, as additional cycles may be necessary before a final answer has been located. Cycle tasks are further made difficult depending upon whether the cycles are *independent* (i.e., the answer identified in one match is not used to carry out a second match, as in the task of listing states with a minimum driving age of 18) or *dependent* (i.e., the answer identified in one match is used in the process of performing a second match, as in the task to identify the state on the bar graph with the lowest crime rate).

Integrate tasks require users to compare or contrast information that has been identified in two or more different locate matches or one or more cycle matches (Kirsch

& Mosenthal, 1992/1993). In general, integrate tasks which require readers to compare information are easier than those that require readers to contrast information. In *generate tasks*, users are required to use prior knowledge (often representing a specific type of content knowledge) to match information in a question or directive to corresponding information in a document (Mosenthal & Kirsch, 1993b). Without the benefit of such knowledge, users often must guess or ask some expert to complete the match.

Matching is also made difficult as the *number of features* required to locate an answer increases (Kirsch & Mosenthal, 1990a). Matches which require the identification of only a single feature are, on average, easier than matches which require the identification of two features; matches requiring two feature matches are easier than three or four feature matches, and so on.

Matching is further made difficult depending upon the *number of responses users must supply* and *whether or not the number of responses, if greater than one, is specified in the question or directive* (Mosenthal & Kirsch, in press). Questions and directives requiring readers to list only one answer are easier than those requiring readers to list two or three answers; questions and directives requiring readers to list two or three answers are easier than those requiring readers to list four answers. Questions and directives which specify the number of multiple responses to be listed are easier than those which do not specify the number of responses explicitly.

Moreover, matching is made difficult to the extent that users have to make *inferences* to match information in the question or directive to information in the document (Mosenthal & Kirsch, 1993b). In such cases, low text-based inferences (or inferences which can be made within the parameters of the information provided in a document) are easier to make than high text-based inferences (or inferences which require some combination of knowledge of the text and specialized prior knowledge).

Similarly, type of match is made difficult to the extent that, once a match has been made, readers then have to choose between two or more pieces of information in order to complete the new (or requested) information frame and, thus, answer the question (Mosenthal & Kirsch, 1991c). In some instances, readers may have to make a low or high text-based inference to determine why one of several possible answers best completes a requested information frame, or readers may have to identify conditional information which renders one answer more consistent with the conditions established by a question or directive.

Based on the preceding observations, the following rules, shown in Figure 3, have been used to score for type of match in this study (Mosenthal & Kirsch, in press). Following Meyer, Marsiske, and Willis (1993), these rules are specified additively. In other words, a task might be assigned a difficulty score of 4 because: It is a literal (add 0) locate (add 1) task which involves a two-feature match (add another 1) that requires readers to list three responses (add another 1) whose actual number is not explicitly specified in the question or directive (add another 1) (type-of-match score total = 4).

Note that, while the scoring system for type of match (as presented in Figure 3) could theoretically generate scores as high as 19 or 20, this was not the case in the DOL and NALS assessments. Rather difficulty scores for type of match ranged from one to eight. The ceiling of eight was not set arbitrarily. Rather, this ceiling (as was the case in prose) reflects the range of difficulty combinations which commonly characterize tasks found in society and the workplace (Painchaud, & Jezak, 1994). While more difficult tasks could be conceived in designing assessments (e.g., a five-feature contrast task requiring high text-based inferencing and ten uncued responses), such tasks would indeed be extremely artificial and would bear little resemblance to those tasks associated with every-day document use (Kirsch & Jungeblut, 1986; Kirsch et al., 1992, 1993).

Insert Figure 3 about here.

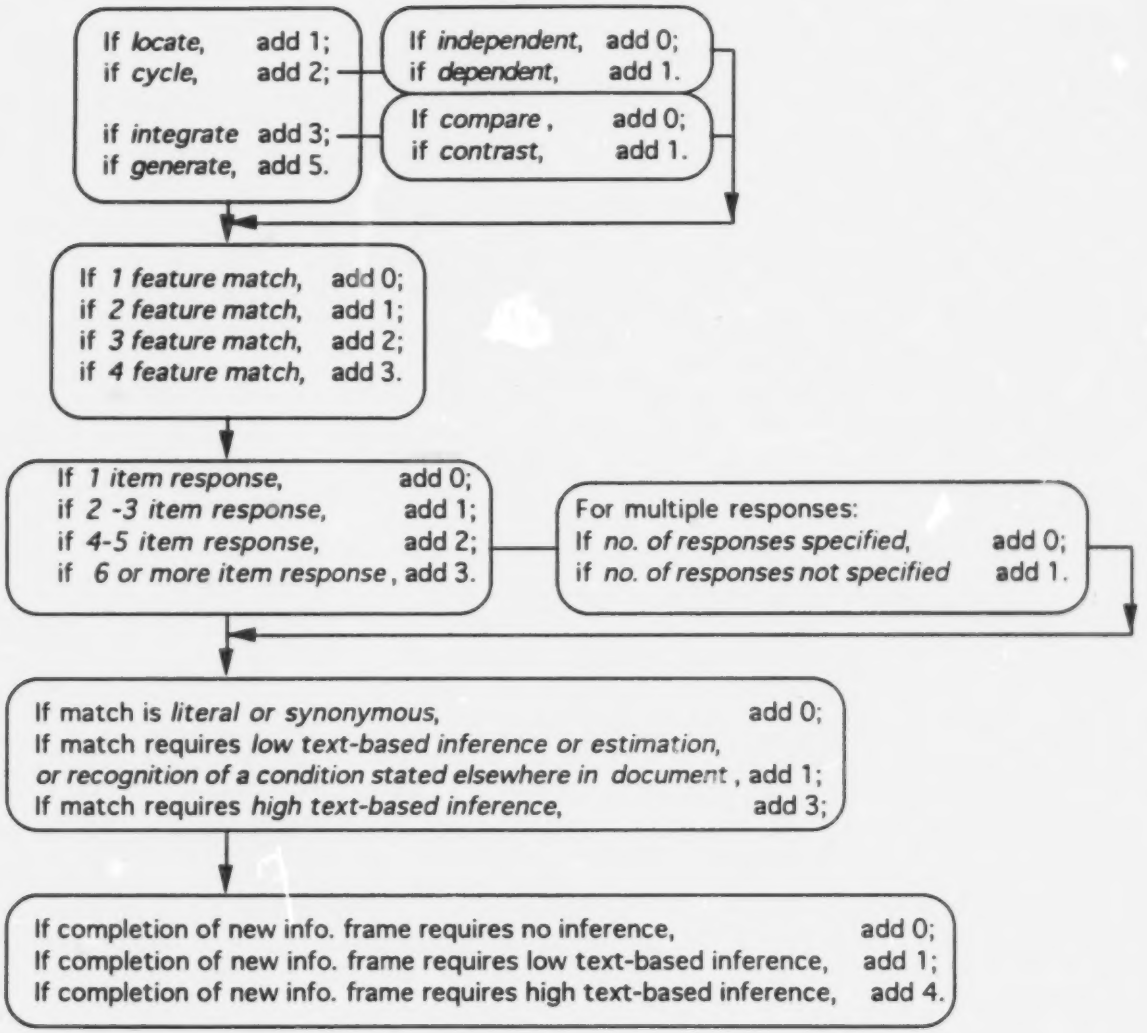

If locate, add 1;
if cycle, add 2;
if integrate add 3;
if generate, add 5.
If independent, add 0;
if dependent, add 1.
If compare, add 0;
if contrast, add 1.
If 1 feature match, add 0;
if 2 feature match, add 1;
if 3 feature match, add 2;
if 4 feature match, add 3.
If 1 item response, add 0;
if 2 -3 item response, add 1;
if 4-5 item response, add 2;
if 6 or more item response, add 3.
For multiple responses:
If no. of responses specified, add 0;
if no. of responses not specified add 1.
If match is literal or synonymous, add 0;
If match requires low text-based inference or estimation,
or recognition of a condition stated elsewhere in document, add 1;
If match requires high text-based inference, add 3;
If completion of new info. frame requires no inference, add 0;
If completion of new info. frame requires low text-based inference, add 1;
If completion of new info. frame requires high text-based inference, add 4.

Figure 3. Additive Scoring Rules for Type of Match in Document Processing

In sum, given the 117 document tasks on the combined DOL and NALS assessments, their type-of-match scores ranged from one (i.e., easiest) to eight (i.e., most difficult). The mean type-of-information score was 2.91 (with a standard deviation of 1.50) with a median score of three. The mean values for type of match by Level were 1.42 for Level 1; 2.47 for Level 2; 3.50 for Level 3; 4.00 for Level 4; and 5.90 for Level 5.

In addition to type of information and type of match, a third process variable is *plausibility of distractors* (Kirsch & Mosenthal, 1990a; Meyer et al., 1993). This variable has to do whether or not there are features from a question or directive's given and/or requested information which appear in the document but, once matched on or identified, do not yield the correct requested information. Based on previous research, Kirsch and Mosenthal (1990a) found that tasks are easiest to process when there are no plausible distractors in a document. (In the current study, such tasks were assigned a score of 1 for plausibility of distractors.) This is often the case when there is only a single item in a list or there is only one list with a unique label unrelated to the other labels in a document.

Tasks become slightly more difficult when there is more than one item in a list in which one is searching for requested information or when there are labels in other lists that bear resemblance in kind to the label on which one is searching. (In this study, such tasks were assigned a score of 2 for plausibility of distractors.)

Tasks become more difficult when one or more features from both given and requested information appear in different matrix cells or in lists other than the cell or list in which an answer actually appears. (In such cases, tasks in this study were assigned a score of 3 for plausibility of distractors.)

Tasks continue to increase in difficulty when one or more features from both requested and given information appear in the same matrix cell or list other than the

answer node. (In such instances, tasks in this study were assigned a score of 4 for plausibility of distractors.)

Finally, tasks are most difficult when one or more features from both requested and given information appear in the same matrix cell or list as the answer. (In this study, such tasks were assigned a score of 5 for plausibility of distractors.)

Given the 117 document tasks in the combined DOL and NALS assessment, their plausibility-of-distractors scores ranged from one (i.e., easiest) to five (i.e., most difficult). The mean plausibility-of-distractors score was 2.69 (with a standard deviation of 1.16) with a median score of two. The mean values for plausibility of distractors were 1.73 for Level 1; 2.28 for Level 2; 3.11 for Level 3; 4.30 for Level 4; and 4.20 for Level 5.

To ensure reliability, two raters independently scored all the tasks comprising the combined DOL and NALS document scales in terms of type of readability, information requested, type of match, and plausibility of distractors. There was 96 percent agreement on readability, 93 percent agreement on type of information, 81 percent agreement on type of match, and 87 percent on plausibility of distractor. Differences between raters were discussed and were agreed upon through consensus.

Defining and Illustrating the Five Levels of Document Proficiency

After each of the variables had been scored in terms of their difficulty, we again looked for patterns of similarity among variables as they were distributed by Levels. These patterns within and between Levels and their related constructs are described below.

Document Level 1. As shown in Table 5, we identified tasks in Level 1 as those which ranged below 225 in RP80 value. The process variables in this range tend to include combination scores of 1, 2 or less, 2 or less. This set of combinations accounted for 21 out of the 25 tasks within this Level (or 84 percent). The average difficulty

value of Tasks at this Level was 198. Twenty-three percent of the adults in the U. S. perform at this level (Kirsch et al., 1993).

Most tasks at this Level require readers to identify information which is quite concrete, including a 'person,' or 'thing,' as well as an 'amount,' 'type of,' 'temporal,' 'action,' or 'location.' Moreover, to complete these tasks, readers must process relatively brief documents to locate a single piece of information which is identical to (or synonymous with) the information given in the question or directive. In some cases, readers must enter personal information (e.g., their name and age) onto a document. If distractors appear in the document, they tend to be representative of either given or new information but not both.

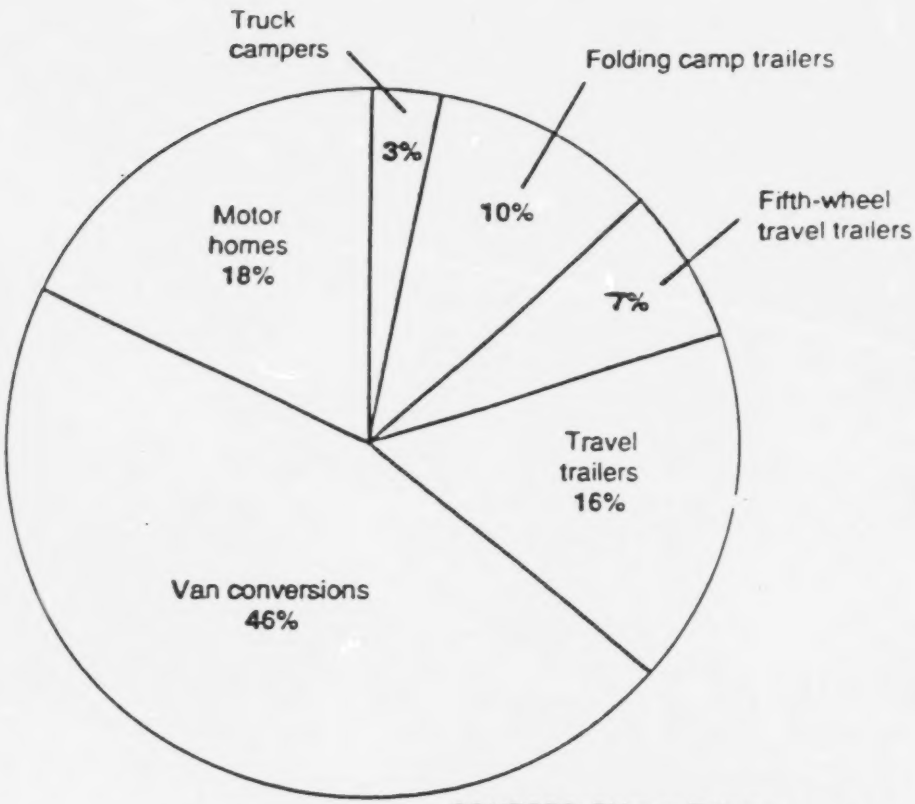
An example of a Level 1 task (with an RP80 of 214) was one which applies to the stimulus F below. This document consists of one label and 12 items and has a readability level of one out of 11. The question based on this pie chart asked, "Which type of recreation vehicle accounted for three percent of the total sales for 1987?" To complete this question, readers have to recognize that the requested information is a "type of." To identify this information, readers must make a literal match between 'three percent' in the question and '3%' in the chart. Note that although there are other recreation vehicles listed besides the answer (i.e., truck campers), there are no percents which include the amount 'three.' Thus, while this task has distractors for requested information, there are no distractors for given information. This contributes to the relative ease of this task.

Insert F about here.

Document Level 2. We identified tasks in Level 2 as those which range between 226 and 275 in RP80 value (see Table 5). The process variables in this range tend to include combination scores which represent a higher difficulty value than those in Level 1. Thus in Level 2, we find process combination scores of: 2, 2, 2; or 3, 3 or less, 2 or

Recreation vehicles

In percent of total sales for 1987



SOURCES: Chicago Tribune, Recreation Vehicles Industry Association

less. These combinations of scores accounted for 42 out of 43 tasks within this Level (or 98 percent). The average difficulty value of tasks at this Level was 250. Twenty-eight percent of the adults in the U. S. perform at this level (Kirsch et al., 1993).

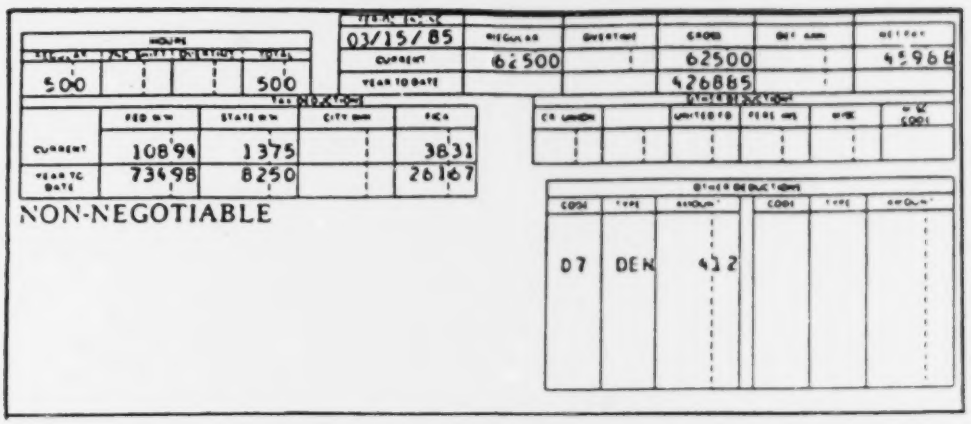
Like tasks in Level 1, most tasks in Level 2 ask readers to complete information which is quite concrete. However, in Level 2, we find some tasks which also require readers to identify 'condition' information. Moreover, tasks at Level 2 often require readers to make a two-feature match or a low level inference to relate given information to information in a document. Other tasks require readers to make two or more dependent cycle matches between a legend and a graph, or between two different parts of a document. In other instances, tasks may require readers to integrate information within a document. Finally, tasks at this Level tend to have a distractor for both given and new information present but not in the same node as the answer.

A task representative of this Level had a difficulty value of 261 and was based on the stimulus G shown below. This document consists of 34 labels and 16 items and has a readability level of five out of 11. This task included the question, "What is the gross pay for this year to date?" To identify this amount, readers must make a two-feature match, identifying both 'gross pay' and 'year to date' correctly in the document. Having made this match, readers then must find the answer '4268.85' in the table. What further contributes to the difficulty of this task is the fact that there are distractors for both given and new information. In searching only on 'gross,' readers may inadvertently select the amount '625.00' which is actually 'current gross,' or they may select any three of the other amounts associated with 'year to date' listed in the table labeled 'Tax Deductions.'

Insert **G** about here.

G

What is the gross pay for this year to date?



					03/15/85				
					CURRENT	62500		62500		45968
					YEAR TO DATE			426885		
500				500					

	FED W/H	STATE W/H	CITY W/H	FICA
CURRENT	10894	1375		3831
YEAR TO DATE	73498	8250		26167

NON-NEGOTIABLE

Document Level 3. We identified tasks in Level 3 as those which range between 276 and 325 in RP80 value. The process variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 2. Thus in Level 3, we find process combination scores of 4 or greater, 3 or less, 3 or less. This combination accounted for 25 out of 28 tasks within this Level (or 89 percent). The average difficulty value of tasks at this Level was 301. Thirty-one percent of the adults in the U. S. perform at this level (Kirsch et al., 1993).

Tasks at this Level often require readers to identify condition information in addition to the usual amount, temporal, and location information. In terms of type of match, Level 3 tasks again require readers to make literal, synonymous, and low level inference matches between the question or directive, and the document. However, unlike Level 1 and 2 locate tasks, Level 3 tasks may require readers to compare or contrast information as well as identify simple patterns or trends . Also the questions and directives of Level 3 tasks tend to require multiple feature matching involving tables which contain nested information. Distracting information for given and requested information tends to be present and often appears in the same node, but not in the node where the answer occurs.

One task at Level 3 (with a difficulty value of 305) involved the stimulus shown in H. This document consists of 46 labels and 353 items and has a readability level of nine out of 11. The question associated with this stimulus was as follows:

> You need to smooth wood in preparation for sealing and plan to buy garnet sandpaper. What type of sandpaper should you buy?

To complete this question, readers must recognize that the requested information is a "type of" sandpaper. To identify this information, readers, at a minimum, must match on 'preparation for sealing' and 'garnet sandpaper' in the question and the corresponding information in the Abrasive Selection Guide, thus making this task a two-feature match. Once this has been completed, readers must identify 'F' as the appropriate abbreviation

for the type of garnet sandpaper needed to prepare wood for sealing. To identify what 'F' stands for, readers must make a dependent cycle involving information in the legend, which reveals that 'F' stands for 'Fine' (i.e., the correct answer to the question).

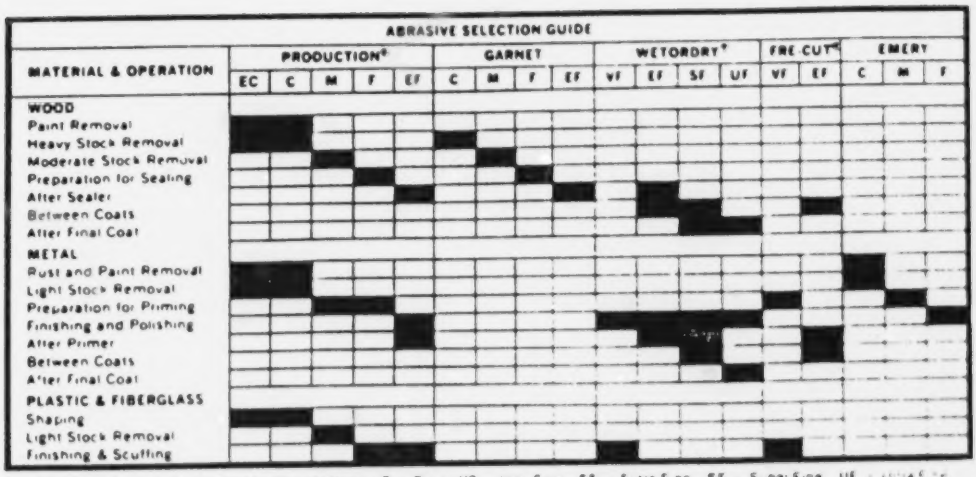
In terms of plausible distractors, there is only one mention of 'Preparation' and 'for Sealing' in the list of operations labeled 'Wood.' While there is mention of 'Preparation' in the list labeled 'Metal,' there is no mention of 'for Sealing.' While there is a type of sandpaper associated with 'Preparation for Sealing' under 'Production,' this type of sandpaper is also 'Fine.' The only other distractors which occur are those that appear in the list of garnet sandpapers to be used in sanding wood. However, a different type is specified for four of the operations and no type is recommended for three of the operations. As such, this task involves distractors for both given and requested information but they appear in different nodes from one another and not in the same node as the answer.

Insert H about here.

Document Level 4. We identified tasks in Level 4 as those which range between 326 and 375 in RP80 value. The process variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 3. Thus, in Level 4, we find process combination scores of 4, 4 or greater, 3 or less. This combination accounted for seven out of ten tasks within this Level (or 70 percent). The average difficulty value of tasks at this Level was 345. Fifteen percent of the adults in the U. S. perform at this level (Kirsch et al., 1993).

Tasks at this Level often require readers to identify rather abstract information, including 'contrast' and 'equivalence.' Level 4 tasks also require readers to make more difficult contrasts and to identify more complex patterns or trends than are characteristic of Level 3 tasks. Also, Level 4 tasks tend to involve multiple feature matching. At the same time, Level 4 tasks invariably include plausible distractors for

H



EC = Extra Coarse C = Coarse M = Medium F = Fine VF = Very Fine EF = Extra Fine SF = Super Fine UF = Ultra Fine

SAFETY INFORMATION:

- Wear approved safety goggles when sanding
- Use particle/dust mask or other means to prevent inhalation of sanding dust
- When using power tools, follow manufacturer's recommended procedures and safety instructions

Reprinted by permission of and copyrighted by the 3M Co.

BEST COPY AVAILABLE

both given and requested information that appear together in the in the same node which, in some cases, may include the answer node.

An example of a Level 4 task was one based on the stimulus shown in I. This document consists of 180 labels and 263 items and has a readability level of ten out of 11. A question associated with this stimulus was as follows:

> On Saturday afternoon, if you miss the 2:35 bus leaving Hancock and Buena Ventura going to Flintridge and Academy, how long will you have to wait for the next bus?

The choice of times associated with this question included: (a) Until 2:57 p.m., (b) Until 3:05 p.m., (c) Until 3:35 p.m., (d) Until 3:57 p.m., and (e) I don't know. Together this question, the distractors, and the related stimulus comprised a task which had a difficulty value of 348.

To answer this question, readers must recognize that the requested information is a time. To identify this time, readers must make a four feature match between 'afternoon,' '2:35,' 'leaving Hancock and Buena Ventura,' and 'going to Flintridge and Academy' in the question, and 'PM,' '2:35,' 'Leave Hancock and Buena Ventura,' and 'Arrive Flintridge and Academy,' respectively. Next, readers must note the condition that the 3:05 bus listed as the next scheduled departure applies 'Monday through Friday only.' Since it is Saturday, readers must avoid the highly plausible '3:05' distractor and proceed by selecting the departure time (i.e., '3:35') that applies to Saturday as well as weekdays.

Insert I about here.

Document Level 5. Finally, we identified tasks in Level 5 as those which are above 375 in RP80 value. The process variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 4. Thus, in Level 5, we find process combination scores of 5, 5 or greater, 5 or less. This

ROUTE **5**

VISTA GRANDE

This bus line operates Monday through Saturday morning local service to most neighborhoods in the northeast section.

Buses run thirty minutes apart during the morning and afternoon rush hours Monday through Friday.

Buses run one hour apart at all other times of day and Saturday.

No Sunday, holiday or night service.

OUTBOUND

from Terminal

	Leave Downtown Terminal	Leave Hancock and Buena Ventura	Leave Casdon	Leave Rustic Hills	Leave North Carefree and Ore Blanco	Arrive Flamingo and Academy
AM	6:20	6:35	6:45	6:50	7:03	7:15
	6:50	**7:05**	**7:15**	**7:20**	**7:33**	**7:45**
	7:20	7:35	7:45	7:50	8:03	8:15
	7:50	**8:05**	**8:15**	**8:20**	**8:33**	**8:45**
	8:20	8:35	8:45	8:50	9:03	9:15
	8:50	**9:05**	**9:15**	**9:20**	**9:33**	**9:45**
	9:20	9:35	9:45	9:50	10:03	10:15
	10:20	10:35	10:45	10:50	11:03	11:15
	11:20	11:35	11:45	11:50	12:03	12:15
PM	12:20	12:35	12:45	12:50	1:03	1:15
	1:20	1:35	1:45	1:50	2:03	2:15
	2:20	2:35	2:45	2:50	3:03	3:15
	2:50	**3:05**	**3:15**	**3:20**	**3:33**	**3:45**
	3:20	3:35	3:45	3:50	4:03	4:15
	3:50	**4:05**	**4:15**	**4:20**	**4:33**	**4:45**
	4:20	4:35	4:45	4:50	5:03	5:15
	4:50	**5:05**	**5:15**	**5:20**	**5:33**	**5:45**
	5:20	5:35	5:45	5:50	6:03	6:15
	5:50	**6:05**	**6:15**	**6:20**	**6:33**	**6:45**
	6:20	6:35	6:45	6:50	7:03	7:15

INBOUND

toward Terminal

Leave Flamingo and Academy	Leave North Carefree and Ore Blanco	Leave Rustic Hills	Leave Casdon	Leave Hancock and Buena Ventura	Arrive Downtown Terminal
6:15	6:27	6:42	6:47	6:57	7:15
6:45	**6:57**	**7:12**	**7:17**	**7:27**	**7:45** Monday through Friday only
7:15	7:27	7:42	7:47	7:57	8:15
7:45	**7:57**	**8:12**	**8:17**	**8:27**	**8:45** Monday through Friday only
8:15	8:27	8:42	8:47	8:57	9:15
8:45	**8:57**	**9:12**	**9:17**	**9:27**	**9:45** Monday through Friday only
9:15	9:27	9:42	9:47	9:57	10:15
9:45	**9:57**	**10:12**	**10:17**	**10:27**	**10:45** Monday through Friday only
10:15	10:27	10:42	10:47	10:57	11:15
11:15	11:27	11:42	11:47	11:57	12:15
12:15	12:27	12:42 p.m.	12:47 p.m.	12:57 p.m.	1:15 p.m.
1:15	1:27	1:42	1:47	1:57	2:15
2:15	2:27	2:42	2:47	2:57	3:15
3:15	3:27	3:42	3:47	3:57	4:15
3:45	**3:57**	**4:12**	**4:17**	**4:27**	**4:45** Monday through Friday only
4:15	4:27	4:42	4:47	4:57	5:15
4:45	**4:57**	**5:12**	**5:17**	**5:27**	**5:45** Monday through Friday only
5:15	5:27	5:42	5:47	5:57	6:15
5:45	5:57	6:12	6:17	6:27	**8:45** Monday through Friday only

Monday through Friday only

You can transfer from this bus to another headed anywhere else in the City bus system.

To the left of a passing number, add the digit of that line's number to the passing time.

BEST COPY AVAILABLE

combination accounted for nine out of ten tasks within this level (or 90 percent). The average difficulty value of tasks at this Level was 399. Three percent of the adults in the U. S. perform at this level (Kirsch et al., 1993).

Tasks at this Level tend to require readers to identify rather abstract information, including 'contrast' and 'patterns.' Level 5 tasks again tend to involve multiple feature matching but also require greater degrees of inferencing in matching information and in identifying requested information. Some of these tasks require readers to provide multiple responses but do not designate how many responses are needed. These tasks also require readers to identify conditional information in identifying requested information. Moreover, Level 5 tasks often include plausible distractors for both given and requested information which appear in the same node as the answer.

An example of a Level 5 task (with an RP80 of 387) was based on the stimulus shown in J. This document consists of nine labels and 141 items and has a readability level of five out of 11. A directive associated with this stimulus was as follows:

> Banks that issue credit cards are organized into two categories in the table. List the two categories. Using information given in the table, describe two differences between these two categories.

To answer this question, readers must recognize that the requested information involves 'contrast'. To complete this type of requested information, readers must first identify the two labels associated with each of the combined lists in the stimulus. One of these combined lists is labeled 'Best Deals for People who Carry Balances,' and the second is labeled 'Best Deals for People Who Pay Off Entire Balance Monthly.' Having identified the two differently labeled combined lists, readers must then integrate information within the columns labeled 'Interest Rate' and 'Annual Fee.' Within the former, readers must identify that interest rates for people who pay off their entire balance monthly are higher than those interest rates for people who carry balances.

Within the latter, readers must identify that there is no annual fee for people who pay off their entire balance monthly while there is an annual fee for people who carry balances (This appears to be the case for institutions except Manufacturers Bank in Wilmington, Delaware). In this task, while the type of information and type of match are quite difficult, plausibility of distractors is rather low, as no other columns of information exist which suggest differences between the two categories of credit card users.

Insert J about here.

Results

Correlations

In the above, four variables were described and illustrated which, based on previous research (Kirsch & Mosenthal, 1990a; Mosenthal & Kirsch, in press), have been shown to influence the difficulty of document tasks. To examine the relations among these variables and task difficulty, we began by computing the intercorrelations between the structure and process variables and task RP80 and Level scores. These correlations are presented in Table 6 below.

Overall, comparing the relations between structure and process variables and task difficulty, the following was found. Type of match correlated highest with RP80 (.81) and Level (.82), followed by plausibility of distractors (.72 with RP80 and .72 with Level) and type of information (.53 with RP80 and .56 with Level). Readability correlated .54 with RP80 and .49 with Level. Among the process variables, there was moderate intercorrelation between type of match and type of information (.49), as well as between type of match and plausibility of distractors (.40). The correlation between plausibility of distractors and type of information was only .20. Between the structure and process variables, correlations ranged from low (.19) between readability and type

J

COMPARING CREDIT CARDS

EDITOR'S NOTE: This chart appears in the Money section of The Times every Monday. It lists the best credit-card deals available nationally during the previous week. The best deals for people who carry balances are the cards with the lowest interest rates, and annual fees will differ. The best deals for people who pay off their balances every month are the cards with no annual fee. Residents of New Jersey and Pennsylvania can obtain more information and application forms by contacting the listed institutions. Rates and fees are subject to change.

Here are the best credit-card deals nationally as of last Tuesday. The rates compare with a national average credit-card rate of 18.21 percent. Rates are for conventional credit cards, not premium cards, and information applies to purchases only. Cash advances frequently are charged interest from the date of transaction. Additional fees may be charged such as for exceeding a credit line, making a payment late, obtaining a cash advance or making an ATM transaction, or if a check is returned. V = variable rate.

BEST DEALS FOR PEOPLE WHO CARRY BALANCES

Institution/Location	Interest Rate	Annual Fee	Interest-Free Days/From
■ Arkansas Federal Svgs Bank Little Rock (501) 224-7283	10.92%V	$30	0
■ Republic S&L Milwaukee, Wisc (414) 257-4240	12.24%V 13.92%	$20 $10	0 0
■ People's Bank Bridgeport, Conn (800) 423-3273	12.50%	$20	25/billing
■ Home Plan S&L Des Moines, Iowa (515) 270-2001	12.90%	$25	25/billing
■ Middletown Savings Bank New York (914) 343-1141	13.75%	$15	25/billing
■ Empire of America Buffalo, NY (800) 843-2443 (U.S.)	13.75%	$18	0
■ Republic National Bank Miami (800) 356-0227 (U.S.)	14.00%V	$22	25/billing
■ Manufacturers Bank Wilmington, Del (302) 366-8487 (U.S.)	14.20%V	$0	0
■ First National Bank Cincinnati (513) 632-4111	14.50%V	$20	25/posting
■ San Francisco Federal S&L (415) 686-5700 (U.S.)	14.50%V	$21	25/billing
■ First Atlanta Bank New Castle, Del (800) 241-7990	14.88%	$24	25/billing
■ Chevy Chase Savings Bank Maryland (800) 367-0669 (U.S.)	14.90%	$20	25/posting
■ Cardinal Federal Svgs Bank Cleveland (800) 423-3236 (U.S.)	15.00%	$18	25/billing
■ Rainier National Bank Seattle (206) 433-7072	15.00%V	$18	25/billing
■ Union National Bank Temple, Okla (800) 351-9125 (U.S.)	15.00%V	$20	0

BEST DEALS FOR PEOPLE WHO PAY OFF ENTIRE BALANCE MONTHLY

Institution/Location	Interest Rate	Annual Fee	Interest-Free Days/From
■ Dauphin Dep. Bank & Trust Harrisburg, Pa (717) 255-2339 (U.S.) (800) 368-2273 (Pa.)	17.90%V	$0	30/billing
■ Security Bank & Trust Southgate, Mich (313) 281-5000	18.00%	$0	25/billing
■ Beverly Bank Chicago (312) 881-2345 (U.S.)	19.80%	$0	25/billing
■ Household Bank Salinas, Cal (800) 223-5279	19.80%	$0	25/billing
■ Imperial Savings San Diego, Cal (800) 542-6209	19.80%	$0	25/billing

SOURCE: BANK RATE MONITOR, N. Palm Beach, Fla. 33408-8888

CARD TIP: No grace-penod cards now appear at the top of the low-rate listings. On those cards, interest is charged on purchases immediately

BEST COPY AVAILABLE

of information to moderate between readability and type of match (.47) and plausibility of distractors (.43).

Insert Table 6 about here.

Regression analyses

Next, two general multiple regression analyses were run using RP80 and Level as measures of task difficulty. Table 7 shows the results of these analyses. The numbers in the table represent the raw beta coefficients for each of the variables included in the regression analyses. In addition, standard errors and *p* -values for each variable are listed. Overall, all three process variables were significant for both RP80 and Level ($p < .01$). In the full regression equation, readability proved nonsignificant ($p > .05$) for both RP80 and Level.

Insert Table 7 about here.

As shown in Table 7, the combined variables accounted for 86 percent of the *R*-squared variance when difficulty was defined using RP80 values; 87 percent of the *R*-squared variance was accounted for when difficulty was defined using Level. When entered into the regression equation by itself, readability was significant ($p = .00$) and accounted for 29 percent of the *R*-squared variance for RP80 and 24 percent of the *R*-squared variance for Level. However, entering this variables in the general regression equation with the three process variables contributed nothing to explained *R*-squared variance for either RP80 or Level. As such, the results of these analyses, as with the prose analyses, again attest to the importance of process variables over structural variables as being the better predictors of task difficulty on the combined DOL and NALS assessments.

Table 6

Intercorrelations between Document Task Difficulty (Represented by RP80 and Level), and Structure and Process Variables for DOL and NALS Combined

	Task difficulty		Structure Variables	Process Variables	
	1	2	3	4	5
Task Difficulty					
1. RP80	-				
2. Level	.95				
Structure Variable		-			
3. Readability	.54	.49			
Process Variables			-		
4. Type of Match	.81	.82	.47	-	
5. Plausibility of Distractors	.72	.72	.43	.40	-
6. Type of Information	.53	.56	.19	.49	.27

Table 7

Raw Beta Coefficients and Standard Errors of Predictive Variables in Regression on Document Task Difficulty Defined Using RP80 and Level for DOL and NALS Combined

	RP80			Level		
	Beta	Std. Error	p^1	Beta	Std. Error	p^1
Structure Variables						
Readability	1.61	.94	.09	.01	.02	.75
Process Variables						
Type of Match	22.46	1.89	.00	.43	.03	.00
Plausibility of Distractors	23.41	2.16	.00	.45	.04	.00
Type of Information	8.51	2.58	.00	.20	.05	.00
Total variance accounted for:						
R^2		86%			87%	
Adjusted R^2		86%			86%	

[1] df = 112

Quantitative Literacy

Since adults are often required to perform numerical operations in every-day life, the ability to perform quantitative tasks is another important aspect of literacy. These abilities may seem, at first glance, to be fundamentally different from the types of skills involved in reading prose and documents and, therefore, appear to extend the concept of literacy beyond its traditional limits. However, research (Kirsch & Jungeblut, 1986, 1992; Kirsch et al., 1993) indicates that the processing of printed information plays an important role in influencing the difficulty of tasks along this scale.

Quantitative Variables

There are several similarities between processing quantitative tasks as there are between processing prose and document tasks. For one, to complete tasks on quantitative, prose, and document scales, readers must use search strategies to match given information in a question or directive to corresponding information in one or more stimuli. Second, the stimuli across tasks represent varying degrees of structural complexity. Third, the tasks of each scale require readers to circumnavigate plausible distractors in the search, identification, and determination of one or more correct responses. And fourth, all three scales require that some form of requested information be identified (Mosenthal & Kirsch, 1993c).

On the other hand, while type of requested information for prose and document literacy varies, the type of information requested in quantitative tasks generally is an *amount* . Moreover, quantitative tasks involve two "formulate" variables not associated with prose and document processing. The first formulate variable is "type of calculation," including such operations as addition, subtraction, multiplication, and division. The second formulate variable is "operation specificity," or the ease by which readers are able to set up a quantitative problem.

In the following section, we begin by discussing structural complexity and process variables as they relate to the characteristics of quantitative tasks summarized in Table 8 below. Next, we describe and illustrate the two formulate variables of type of calculation and operation specificity.

Structural complexity. The stimuli on the quantitative scale had a mean document readability of 3.58 (with a standard deviation of 1.93) and a median of three. The mean readability scores by Level were as follows: 2.00 for Level 1; 3.40 for Level 2; 3.58 for Level 3; 3.38 for Level 4; and 4.56 for Level 5.

Insert Table 8 about here.

Process variables. In addition to determining the readability of the documents on the quantitative scale, we used the same document rules for scoring type of match and plausibility of distractors for this scale. Given the 71 quantitative tasks in the combined DOL and NALS assessments, their type-of-match scores ranged from one (i.e., easiest) to 5 (i.e., most difficult). The mean type-of-match score was 2.48 (with a standard deviation of 1.03) with a median score of 2. The mean values for type of match were 1.00 for Level 1; 2.00 for Level 2; 2.48 for Level 3; 2.58 for Level 4; and 2.78 for Level 5. In short, unlike the broad range of variance for type of match found on the document scale, the range for type of match found on the quantitative scale had less variance.

The plausibility-of-distractors scores for the quantitative tasks again ranged from one (i.e., easiest) to 5 (i.e., most difficult). The mean plausibility-of-distractors score was 2.63 (with a standard deviation of 1.12) with a median score of 2. The mean values for plausibility of distractors were 1.00 for Level 1; 1.60 for Level 2; 2.45 for Level 3; 2.83 for Level 4; and 3.67 for Level 5.

Formulate variables. The formulate variable, type of calculation, involves the various operations that readers use to relate one set of numbers to another in order to

Table 8

Characteristics of Variables (i..e, Readability ('Read'), Type of Match ('TOM'), Plausibility of Distractors ('POD'), Type of Calculation ('TOC') and Operation Specificity ('Op. Sp.') for DOL and NALS Quantitative Tasks by RP80 and Level Difficulty

Description	RP80	Level	Read	TOM	POD	TOC	Op. Sp.
Automatic Teller Deposit Slip	191	1	2	1	1	1	1
Deposit Slip*	220	1	2	1	1	1	1
Receipt for Certified Mail	238	2	3	1	2	1	2
Theater Trip Notice	246	2	2	2	1	2	2
Salt River Recreation Ad	250	2	5	2	2	2	3
Catalog Order Form*	270	2	6	3	2	1	3
Tempra Coupon	273	2	1	2	1	5	4
Check Ledger Entries*	277	3	3	2	2	1	4
Insurance Protection Workform	277	3	2	3	2	3	2
Dupont Carpet Advertisement	278	3	2	3	1	2	4
Salt River Recreation Ad	278	3	5	3	4	1	3
Middle-Class Growth Chart	279	3	4	4	2	2	4
Pest Control Caution	281	3	2	3	3	1	4
Check Ledger Entries*	281	3	3	2	2	2	4
Check Ledger Entries*	282	3	3	1	2	2	4
Money Rates	283	3	4	3	2	2	4
Check Ledger Entries*	283	3	3	2	2	2	4
Ashland Oil Bill	284	3	3	2	4	2	2
Recreation Vehicle Pie Chart	287	3	2	2	2	1	2
Burning Out of Control	288	3	2	3	2	2	4
Dessert Recipes	291	3	5	3	2	2	4
Denver/Minneapolis Sched.*	295	3	3	3	3	5	3
LPGA Money Leaders	296	3	5	2	2	2	4
Comparing Credit Cards	300	3	4	1	2	2	4
Businessland Printer Stand	302	3	2	3	3	2	4
Consumer Report Books	303	3	7	2	3	1	4
Wall Panel Diagram + Materials	306	3	5	1	2	4	3
Valet Airport Parking Discount	307	3	2	3	3	2	4
Ashland Oil Bill	308	3	3	2	2	3	5
Unit Prices for Peanut Butters*	311	3	2	2	3	4	6
Money Rates	312	3	4	3	3	2	2
Valet Airport Parking Discount	315	3	2	2	3	3	4
Pizza Coupons	316	3	2	3	3	1	4
LPGA Money Leaders	321	3	5	1	2	4	3
South Highland Bus Schedule	321	3	9	4	2	2	5
Dessert Recipes	322	3	5	3	2	3	6
Tempra Dosage Chart	322	3	5	3	3	2	4
Denver/Minneapolis Sched.*	325	3	3	3	3	5	3
Welcome to the Desert Highway	330	4	2	3	1	5	6
Lunch Menu*	331	4	2	2	2	5	4
Butcher Captures Iditarod	332	4	2	3	2	4	6
Salt River Recreation Tube Ride	333	4	5	2	4	1	5
Businessland Printer Stand	339	4	2	3	4	3	5
Recreation Vehicles	342	4	2	3	4	4	3
Thrift Grocery Store Coupons	343	4	2	4	1	3	6
Washington Train Schedule	343	4	7	4	4	2	5
Red Bud Room Setup*	345	4	3	4	4	1	5
Lunch Menu*	348	4	2	1	2	5	7
Washington Train Schedule	348	4	7	4	4	2	5
Cost to Raise a Kid	350	4	2	1	2	3	5
El Paso Gas & Electric	350	4	8	3	4	2	2
Spotlight Economy	354	4	8	5	4	2	2
Recreation Vehicles	355	4	2	1	2	3	6

BEST COPY AVAILABLE

Unit Prices for Peanut Butters*	355	4	2	1	2	4	5
Insurance Protection Workform	355	4	2	1	2	5	4
Effects of High Blood Pressure	360	4	2	1	2	3	6
Camp Advertisement	364	4	2	2	4	5	4
Middle-Class Growth Chart	365	4	4	3	2	4	4
Camp Advertisement	366	4	2	3	4	5	4
Are You Eligible for SSI?	368	4	2	2	4	3	6
Companies Share Cig. Market	370	4	6	3	2	3	6
Automobile Maintenance Record	375	4	3	3	2	5	4
Tempra Dosage Chart	379	4	5	3	5	3	5
Catalog Order Form*	382	5	6	4	5	5	5
How Education Dollars Grow	384	5	6	4	5	2	4
Wall Panel Diagram + Materials	388	5	5	2	3	5	7
Butcher Captures Iditarod	405	5	2	2	1	5	5
Registration Information	406	5	8	2	5	5	5
Dupont Carpet Advertisement	421	5	2	1	2	5	7
Home Equity Loan*	433	5	2	5	5	5	7
Wall Panel Diagram + Materials	436	5	5	2	2	5	7
*Tasks on both DOL and NALS Assessments							

BEST COPY AVAILABLE

produce a sum, difference, product, or quotient. Tasks which involve a single addition tend to be easiest (and, in our analyses, received a score of 1). Tasks which involve a single subtraction tend to be the next easiest (and received a score of 2). Tasks which involve a single multiplication tend to be slightly more difficult (and received a score of 3). Tasks which involve a single division tend to be even more difficult (and received a score of 4). Finally, tasks which involve two or more operations (such as a division followed by a multiplication) tend to be the most difficult (and received a score of 5 for type of calculation in our analyses).

Given the 71 quantitative tasks in the combined DOL and NALS assessments, their type-of-calculation scores ranged from one (i.e., easiest) to five (i.e., most difficult). The mean type-of-calculation score was 2.92 (with a standard deviation of 1.44) with a median score of three. The mean values for type of calculation by Level were 1.00 for Level 1; 2.20 for Level 2; 2.32 for Level 3; 3.42 for Level 4; and 4.44 for Level 5.

The second formulate variable is operation specificity. This variable involves the process of setting up an arithmetic equation based on the operation parameters established in a task. Moreover, this variable includes the process of identifying and (in some cases) entering numbers that go into an equation. Operation specificity is influenced by the following conditions (see Figure 4 below for a summary). Tasks are easier when numbers to be operated on appear in a row or column format and when they are adjacent to one another. Tasks tend to be more difficult when numbers to be operated on are not in column or row format and when they are nonadjacent.

Tasks are easier if amounts to be operated on do not require search. Tasks become more difficult when amounts have to be identified with search, and labels associated with amounts have to be inferred.

Tasks are easier if an operation is explicitly signaled by '+,' '-,' 'x,' '/,' or if they include such terms as 'add,' 'subtract,' 'multiply,' and 'divide.' Tasks become more difficult if an operation is specified by an explicit semantic relation statement, such as

'how much more,' 'how much less,' 'how many times,' and 'calculate the difference.' Tasks become even more difficult if an operation is specified indirectly using such terms as 'how much is saved,' 'what is the deduction,' and 'what is the net profit.' Finally, tasks are even more difficult if readers must use a unit ratio (e.g., 'miles per gallon,' 'cost per square foot,' 'square yards') stated in a question or directive to formulate an equation and determine an operation.

Overall, tasks are easier when they involve the manipulation of numbers identified in a current task; tasks become more difficult when they involve the manipulation of numbers which are the outcome of operations in preceding tasks.

Finally, tasks are easier when amounts require no transformation; tasks become more difficult when amounts do require transformations (e.g., transforming two times in different units of hours and minutes, or transforming numerical ratios with different unit denominators).

Insert Figure 4 about here.

Note that, as with the variable type of match, operation specificity is scored additively (as shown in Figure 4). Although this variable could range as high as nine, the highest operation specificity score for the combined 71 quantitative tasks was seven (see Table 8). The lowest score for operation specificity was one. The mean operation-specificity score was 4.23 (with a standard deviation of 1.46) and a median score of four. The mean values for operation specificity were 1.00 for Level 1; 2.80 for Level 2; 3.77 for Level 3; 4.79 for Level 4; and 5.78 for Level 5.

To ensure reliability, two raters independently scored all 71 tasks on the combined DOL and NALS quantitative scale in terms of type of readability, type of match, plausibility of distractors, type of calculation, and operation specificity. There was 91 percent agreement on readability, 87 percent agreement on type of match, 85 percent on plausibility of distractors, 94 percent on type of calculation, and 89 percent on

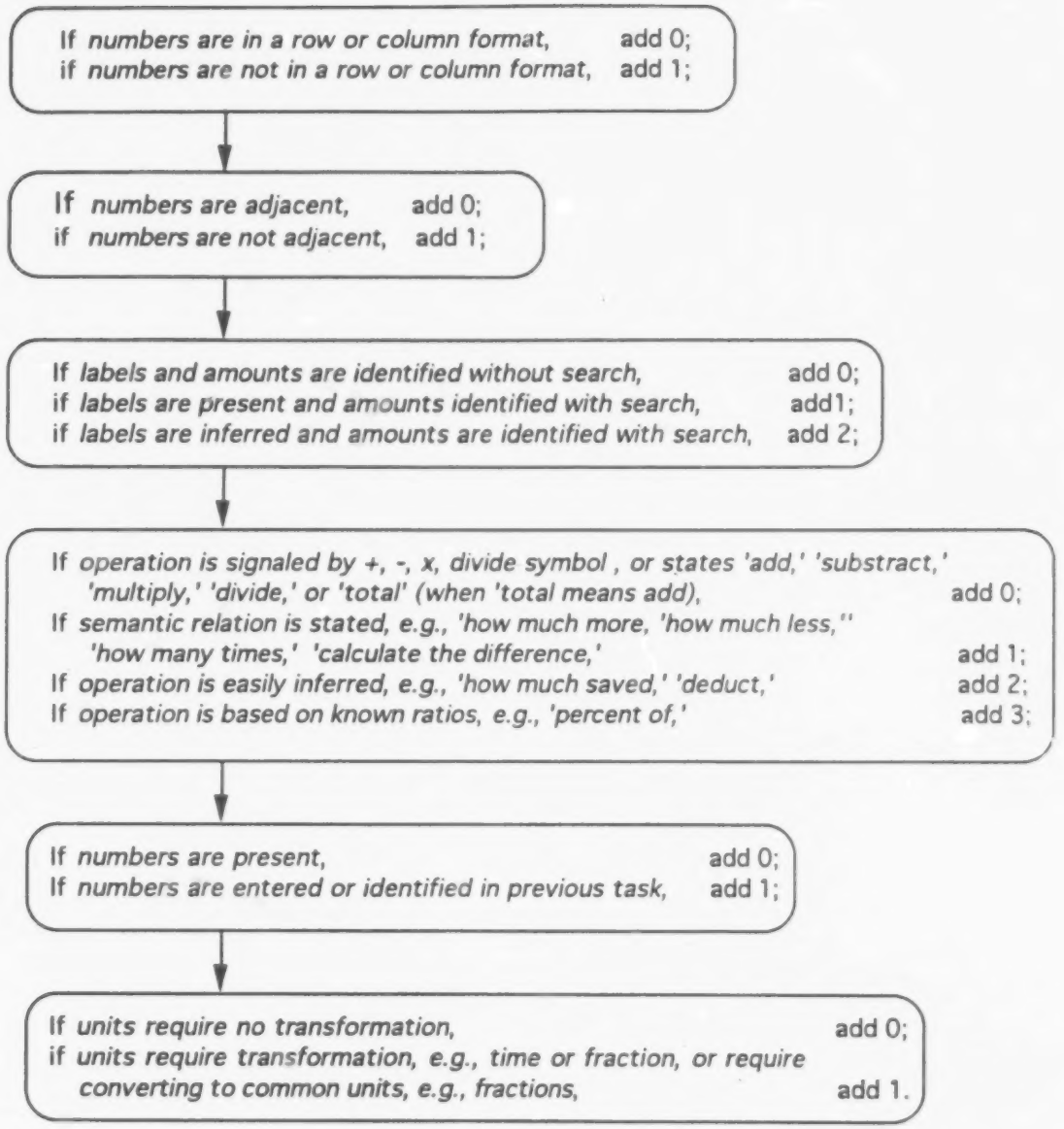

If numbers are in a row or column format, add 0;
if numbers are not in a row or column format, add 1;
If numbers are adjacent, add 0;
if numbers are not adjacent, add 1;
If labels and amounts are identified without search, add 0;
if labels are present and amounts identified with search, add1;
if labels are inferred and amounts are identified with search, add 2;
If operation is signaled by +, -, x, divide symbol , or states 'add,' 'substract,' 'multiply,' 'divide,' or 'total' (when 'total means add), add 0;
If semantic relation is stated, e.g., 'how much more, 'how much less,'' 'how many times,' 'calculate the difference,' add 1;
If operation is easily inferred, e.g., 'how much saved,' 'deduct,' add 2;
If operation is based on known ratios, e.g., 'percent of,' add 3;
If numbers are present, add 0;
If numbers are entered or identified in previous task, add 1;
If units require no transformation, add 0;
if units require transformation, e.g., time or fraction, or require converting to common units, e.g., fractions, add 1.

Figure 4. Additive Scoring Rules for Operation Specificity

operation specificity. Differences between raters were discussed and were agreed upon through consensus.

Defining and Illustrating the Five Levels of Quantitative Proficiency

After each of the variables had been scored in terms of their difficulty, we again looked for patterns of similarity among variables within and between Levels. As with the prose and document scales, we defined the five Levels on the quantitative scale in terms of 50-point increments based on RP80 values. These patterns of similarity and their related constructs by 50-point increments are described below.

Quantitative Level 1. As shown in Table 8, tasks in Level 1 range below 225. The plausibility distractor variable and the two formulate variables in this range tend to include combination scores of 1, 1, and 2 or less. This combination accounted for two out of the two tasks within this Level (or 100 percent). These tasks had an average RP80 value of 206. Twenty-two percent of the adults in the U. S. perform at this Level (Kirsch et al., 1993).

Tasks at this Level require readers to perform a single, relatively simple arithmetic operation, which usually is addition. The numbers to be operated on tend to be adjacent to one another, appearing in the same row or column. The label(s) associated with these numbers are provided and the numbers to be operated on usually can be identified with little or no search. In most instances, the type of operation is signaled by an arithmetic sign (e.g., '+' or '-') or the term 'total' (when used to mean 'addition'). The numbers to be operated on are unrelated to previous tasks and require no transformations. In most cases, there are no other amounts listed which could serve as plausible distractors.

An example of a Level 1 task was based on the stimulus K below. This document consists of 13 labels and 18 items and has a readability level of two out of 11. The directive related to this document instructed readers as follows:

You wish to use the automatic teller machine at your bank to make a deposit. Figure the total amount of the two checks being deposited. Enter the amount on the form in the space next to TOTAL.

The task associated with this directive and stimulus **K** had a difficulty value of 191.

To complete this task, readers have to recognize that 'TOTAL' in this case means 'add.' Moreover, readers must identify the two check amounts (i.e., '$557.19' and '$75.00'), which are adjacent and in column format. Once having added the two check amounts, readers then must enter the sum (i.e., '$632.10') in the appropriate space. Note that, since no other amounts are listed in the document, there are no distractors present in this task.

Insert **K** about here.

Quantitative Level 2. We identified tasks in Level 2 as those which range between 226 and 275 in RP80 value (see Table 8). The formulate variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 1. Thus, in Level 2, we find formulate combination scores of 2, 2 or 3, 2 or less; or 3, 1,1. These combinations accounted for four out of five tasks within this Level (or 80 percent). These tasks had an average RP80 value of 255. Twenty-five percent of the adults in the U. S. perform at this Level with 80 percent probability (Kirsch et al., 1993).

Like tasks in Level 1, some tasks in Level 2 require readers to perform a simple addition. However, in Level 2, we also find tasks which require readers to perform a simple subtraction. As with Level 1 tasks, the numbers to be operated on in Level 2 tasks tend to be adjacent to one another, appearing in the same row or column. The label(s) associated with these numbers are provided. However, the numbers to be operated on usually require some search involving simple cycling. In some instances, operations are no longer signaled by arithmetic signs but rather in terms of relational

K

Availability of Deposits

Funds from deposits may not be available for immediate withdrawal. Please refer to your institution's rules governing funds availability for details.

Crediting of deposits and payments is subject to verification and collection of actual amounts deposited or paid in accordance with the rules and regulations of your financial institution.

PLEASE PRINT

YOUR MAC CARD NUMBER (No PINs PLEASE)
111 222 3334

YOUR FINANCIAL INSTITUTION
Union Bank

YOUR ACCOUNT NUMBER
987 555 674

YOUR NAME
Chris Jones

CHECK ONE ☐ DEPOSIT or ☐ PAYMENT

CASH	$	00
LIST CHECKS BY BANK NO	ENDORSE WITH NAME & ACCOUNT NUMBER	
	557	19
	75	00
TOTAL		

DO NOT DETACH TICKET

DO NOT FOLD NO COINS OR PAPER CLIPS PLEASE

BEST COPY AVAILABLE

statements such as 'how much more' and 'how much less.'. The numbers to be operated on again are unrelated to previous tasks and require no transformations. However, in some instances, there are other numbers present which serve as simple distractors.

A task representative of this Level had a difficulty value of 246 and was based on the stimulus L below. This document consists of six labels and 15 items and has a readability level of two out of 11. This task included the question, "The price of one ticket and bus for 'Sleuth' costs how much less than the price of one ticket and bus for 'On the Town'?" To answer this question, readers must recognize that 'how much less' signals subtraction. To identify the cost of a ticket and bus for the two shows respectively, readers must cycle, identifying the adjacent amounts of '$11.00' and '$8.50' in the same column. Subtracting the latter from the former, readers arrive at the difference of '$2.50.'

Insert L about here.

Quantitative Level 3. We identified tasks in Level 3 as those which range between 276 and 325 in RP80 value. The formulate variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 2. Thus in Level 3, we find plausibility of distractors and formulate variable combination scores of 4 or higher, 3 or less, 3 or less. This combination accounted for 26 out of 31 tasks at this Level (or 84 percent). These tasks had an average RP80 value of 294. Thirty-one percent of the adults in the U. S. perform at this Level with 80 percent probability (Kirsch et al., 1993).

Besides requiring readers to perform simple addition and subtraction, tasks at Level 3 begin to include multiplication and division, as well as simple multiple operations (e.g., an addition followed by a subtraction). Unlike in Level 1 and 2 tasks, the numbers to be operated on in Level 3 tasks tend not to be adjacent to one another, although they continue to appear in the same row or column. Again, as in Level 1 and 2

THEATER TRIP

A charter bus will leave from the bus stop (near the Conference Center) at 4 p.m., giving you plenty of time for dinner in New York. Return trip will start from West 45th Street directly following the plays. Both theaters are on West 45th Street. Allow about 1½ hours for the return trip.

Time:	4 p.m., Saturday, November 20		
Price:	"On the Town"	Ticket and bus	$11.00
	"Sleuth"	Ticket and bus	$8.50
Limit:	Two tickets per person		

tasks, the label(s) associated with these numbers in Level 3 tasks are provided and, as in Level 2 tasks, the numbers to be operated on usually require search using cycling. However, unlike Level 1 and 2 tasks, Level 3 tasks usually require readers to infer operations based on accounting terms (e.g., 'how much saved'). The numbers to be operated on again are unrelated to previous tasks. Moreover, some Level 3 tasks require transformations of times and ratios. In most instances, there are other numbers present which serve as plausible distractors.

A task representative of this Level had a difficulty value of 321 and was based on the stimulus **M** below. This document consists of 18 labels and 223 items and has a readability level of nine out of 11. This task included the question:

> Suppose that you took the 12:45 bus from U. A. L. R. Student Union to 17th and Main on a Saturday. According to the schedule, how many minutes is the bus ride?

To complete this question, readers must infer that the time traveled between two points requires subtracting the earlier from the later time. To identify the times, readers must first make a three-feature match based on the features 'p.m.,' '12:45,' and 'bus from U. A. L. R. Student Union' in the question and the corresponding information in the document. Next, readers must cycle and locate the time associated with the labels 'P.M.,' 'Bus arrives at 17th & Main,' and the item '12.45' in the document. Note that, in this case, the time '1:06,' while in the same row as '12:45,' is not adjacent to this time. To subtract '12:45' from '1:06' requires transforming '1:06' to the time '12:66' in order to perform the subtraction and arrive at the answer '21 minutes.' Finally, note that other times are present as distractors but none of which are identical to '12:45' in the column labeled 'Bus Leaves from U. A. L. R. Student Union.'

Insert **M** about here.

M

16 South Highland 16

4	5	6	7
BUS LEAVES from UALR Student Union	Bus arrives at 20th & Woodrow	Bus arrives at 17th & Main	BUS ENDS at Capitol & Louisiana

WEEKDAYS

AM	♿	5:38	5:51	6:00	6:09
		6:11	6:25	6:35	6:45
	♿	6:41	6:55	7:05	7:15
		7:11	7:25	7:35	7:45
	♿	7:41	7:55	8:05	8:15
		8:11	8:25	8:35	8:45
	♿	8:41	8:55	9:05	9:15
		9:14	9:27	9:36	9:45
	♿	9:44	9:57	10:06	10:15
		10:14	10:27	10:36	10:45
	♿	10:44	10:57	11:06	11:15
		11:14	11:27	11:36	11:45
	♿	11:44	11:57	12:06	12:15
PM	♿	12:14	12:27	12:36	12:45
	♿	12:44	12:57	1:06	1:15
		1:14	1:27	1:36	1:45
	♿	1:44	1:57	2:06	2:15
		2:14	2:27	2:36	2:45
	♿	2:44	2:57	3:06	3:15
		3:14	3:27	3:36	3:45
	♿	3:43	3:56	4:05	4:15
		4:13	4:26	4:35	4:45
	♿	4:43	4:56	5:05	5:15
		5:13	5:26	5:35	5:45
	♿	5:45	5:58	6:07	6:17
		6:11	6:22	6:30	-
	♿	6:46	6:57	7:05	-

SATURDAY

AM	♿	5:38	5:51	6:00	6:09
	♿	6:45	6:57	7:06	7:15
	♿	7:45	7:57	8:06	8:15
	♿	8:45	8:57	9:06	9:15
	♿	9:45	9:57	10:06	10:15
	♿	10:45	10:57	11:06	11:15
	♿	11:45	11:57	12:06	12:15
PM	♿	12:45	12:57	1:06	1:15
	♿	1:45	1:57	2:06	2:15
	♿	2:45	2:57	3:06	3:15
	♿	3:45	3:57	4:06	4:15
	♿	4:45	4:57	5:06	5:15
	♿	5:45	5:57	6:06	6:15
	♿	6:44	6:56	7:05	

Reduced from original copy

BEST COPY AVAILABLE

Quantitative Level 4. We identified tasks in Level 4 as those which range between 326 and 375 in RP80 value. The plausibility of distractors and formulate variables in this range include combination scores of 4 or higher, 4, 4 or less; or 6, 4 or less, 4 or less. This combination of scores accounted for 19 out of 24 tasks at this Level (or 79 percent). These tasks had an average RP80 value of 349. Seventeen percent of the adults in the U. S. perform at this Level with 80 percent probability (Kirsch et al., 1993).

Most tasks at Level 4 require readers to perform calculations using multiplication and division, as well as more complex multiple operations (e.g., a multiplication followed by a subtraction). Numbers to be operated on in Level 4 tasks tend not to be adjacent to one another nor do they appear in the same row or column. In some instances, the label(s) associated with these numbers are not provided. In other instances, the labels for numbers are provided but readers must infer equations and operations based upon ratios inferred from statements in the document. In some instances, the numbers to be operated on are related to previous tasks. Moreover, some Level 4 tasks, like Level 3 tasks, may require transformations of times and ratios. And like Level 3 tasks, there usually are other numbers present which serve as distractors. Some of these distractors may involve the situation where distractors for both given and requested information appear in the same node but not in the answer node.

One task representative of this Level had a difficulty value of 355 and was based on the stimulus N below. This document consists of six labels and six items and has a readability level of two out of 11. This task included the directive:

> Estimate the cost per ounce of the creamy peanut butter. Write your estimate on the line provided.

To complete this directive, readers must first identify the unit price label for creamy peanut butter. Next, using the ratio of 'cost per ounce,' they must set up the equation that "if 20 ounces of creamy peanut butter cost $1.99 then one ounce of creamy

peanut butter would cost 'x.'" Solving this equation, readers may opt to divide $1.99 by 20 ounces to arrive at the answer of '10 cents per ounce.' (Note, of course, a similar solution could be obtained if one divided the unit price of '$1.59 per pound' by '16 ounces per pound'; in this case, the unit 'pounds' would cancel one another leaving 'per ounce.') Again, note that plausible distractors appear in this task, as other costs and amounts are listed in the stimulus.

Insert **N** about here.

Quantitative Level 5. Finally, we identified tasks in Level 5 as those which are above 375 in RP80 value. The formulate variables in this range tended to include combination scores which represent a higher difficulty value than those in Level 4. Thus, in Level 5, we find formulate combination scores 5, 5 or higher, 5 or less. This combination accounted for process value characteristics of eight out of nine tasks at this Level (or 89 percent). These tasks had an average RP80 value of 411. Four percent of the adults in the U. S. perform at this Level with 80 percent probability (Kirsch et al., 1993).

All tasks at Level 5 require readers to perform calculations involving multiple operations (which usually include multiplication and division). Numbers to be operated on in Level 5 tasks tend not to be adjacent to one another, nor do they appear in the same row or column. In some instances, the label(s) associated with these numbers are not provided. In other instances, the labels for numbers are provided but readers must infer equations and operations based upon known ratios provided in the document. Moreover, some Level 5 tasks, like Level 3 and 4 tasks, may require transformations of times and ratios. Finally, there are other numbers usually present which serve as distractors. Some of these distractors may appear in the same node as the answer.

N

Unit price	You pay
11.8¢ per oz.	1.89
rich chnky pnt bt	
10693	16 oz.

Unit price	You pay
1.59 per lb.	1.99
creamy pnt butter	
10732	20 oz.

One task representative of this Level had a difficulty value of 421 and was based on the stimulus O below. This document consists of one label and nine items and has a readability level of two out of 11. This task included the directive:

> Suppose that you want to carpet your living room which is 9 feet by 12 feet, and you purchase DuPont Stainmaster carpet at the sale price. Using the calculator, compute the total cost, excluding tax and labor, of exactly enough carpet to cover your living room floor.

To complete this directive, readers must first identify the cost per unit yardage. This requires a simple locate match to identify '$9.49 per sq. yd.' Next, readers must compute the square yardage of the living room. To accomplish this, readers need to transform '9 feet by 12 feet' into yards by dividing each unit of feet by three (i.e., there are three feet per yard). This produces '3 yards by 4 yards.' Multiplying length by width results in '12 square yards.' To determine the cost of covering the living room floor, readers must then multiply '12 square yards' by '$9.49 per square yard.' This process cancels the unit 'per square yard,' leaving the cost '$113.88.'

Note that, similar to most tasks at Level 5, this task involves difficult multiple operations (i.e., a division and two multiplications). Moreover, the numbers to be operated on are not in row or column format nor are they adjacent. The label 'sale price' needs to be inferred and the cost identified with search. The unit 'feet' need to be transformed into 'yards' in order to compute 'cost per square yard.' Finally, operations are based on a knowledge of ratios. In the document, the regular cost of carpet serves as a plausible distractor.

Insert O about here.

The Carpet Store

DUPONT STAINMASTER CARPET

41% OFF DUPONT STAINMASTER

Incredible savings on carefree DuPont carpet.
Your choice of the many luscious colors.
Reg. 15.99

9.49 SQ YD
WITH PADDING

ONE WEEK ONLY!
SAVE ON EVERY STYLE
EVERY COLOR
STAINMASTER

Results

Correlations

In the above, two formulate variables and a strategy variable (i.e., plausibility of distractors) were described which, based on previous research (Mosenthal & Kirsch, 1993c), have been shown to influence the difficulty of quantitative tasks. To examine the relations among these variables and task difficulty, we first computed intercorrelations between these variables and task RP80 and Level scores, as well as between structure and type of match and task difficulty scores. These correlations are presented in Table 9 below.

Overall, comparing the relations between structure, process, formulate variables, and task difficulty, we found the following. Operation specificity correlated highest with RP80 (.80) and Level (.80), followed by type of calculation (.73 with RP80 and .69 with Level). Plausibility of distractors correlated moderately with RP80 (i.e., .37) and Level (i.e., .36). Type of match correlated rather low with RP80 (.23) and Level (.24). Readability correlated very little with RP80 (-.01) and Level (-.05). There was a relatively high intercorrelation between the two formulate variables (.55). Plausibility of distractors correlated .22 with operation specificity and .16 with type of calculation .

Insert Table 9 about here.

Regression analyses

Next, two general multiple regression analyses were run using RP80 and Level as dependent variables. Note that, since readability and type of match were not significant in the correlation analyses, they were not included in these regressions. Table 10 shows the results of the regressions using plausibility of distractors, type of calculation, and operation specificity. As in previous analyses, the numbers in the table

Table 9

Intercorrelations between Quantitative Task Difficulty (Represented by RP80 and Level), and Structure, Process, and Formulate Variables for DOL and NALS Combined

	Task difficulty		Structure Variables	Process Variables		Formu -late
	1	2	3	4	5	7
Task Difficulty						
1. RP80	-					
2. Level	.94					
Structure Variable		-				
3. Readability	.17	.16				
Process Variables			-			
4. Type of Match	.21	.25	.35	-		
5. Plausibility of Distractors	.45	.49	.37	.45	-	
Formulate Variables						
6. Calculate	.63	.55	-.20	-.09	.03	-
7. Operation Specificity	.68	.67	-.04	.06	.09	.46

BEST COPY AVAILABLE

represent the raw beta coefficients for each of the variables. In addition, standard errors and p -values for each variable are listed. Overall, both formulate variables and plausibility of distractors were significant for both RP80 and Level ($p < .01$).

Insert Table 10 about here.

As shown in Table 10, the combined variables accounted for 75 percent of the R-squared variance when difficulty was defined using RP80 values; 71 percent of the R-squared variance was accounted for when difficulty was defined using Level. As such, the results of this analysis attest to the importance that both formulate variables and plausibility of distractors play as predictors of task difficulty on the combined DOL and NALS quantitative scales.

Discussion

Validation of Constructs underlying Proficiency Levels

One purpose of this chapter was to identify and validate the constructs which contribute to the difficulty of tasks on the combined DOL and NALS prose, document, and quantitative scales. In attempting to accomplish this, we identified several variables which significantly accounted for task difficulty as defined using the dependent measures of RP80 values and Level scores. The variables that we examined included structure, strategy (or process), and formulate variables. When comparing structure and strategy variables, the best predictors of prose and document task difficulty for RP80 and Level (see Tables 11 and 12 below, respectively) were strategy variables, i.e., type of match, plausibility of distractors, and type of information. When comparing structure, strategy, and formulate variables, the best predictors of quantitative task difficulty for RP80 and Level were formulate variables, i.e., type of calculation and operation specificity, followed by the strategy variable, plausibility of distractors (see Table 13 below).

Table 10

Raw Beta Coefficients and Standard Errors of Predictive Variables in Regression on Quantitative Task Difficulty Defined Using RP80 and Level for DOL and NALS Combined

Process Variable	Beta	RP80 Std. Error	p^1	Beta	Level Std. Error	p^1
Plausibility of Distractors	17.02	2.66	.00	.04	.05	.00
Formulate Variables						
Operation Specificity	15.51	2.31	.00	.20	.05	.00
Calculate	13.53	2.32	.00	.30	.05	.00
Total variance accounted for:						
R^2		75%			71%	
Adjusted R^2		74%			70%	

[1] df = 67

The three strategy variables accounted for 80 percent of the *R*-squared variance for prose task difficulty and 86 percent of the *R*-squared variance for document task difficulty using RP80 as the dependent measure. The three strategy variables accounted for 80 percent of the *R*-squared variance for prose task difficulty and 87 percent of the *R*-squared variance for document task difficulty using Level as the dependent measure (see Tables 11 and 12, respectively). The two formulate variables and the strategy variable, plausibility of distractors, accounted for 75 percent of the *R*-squared variance for quantitative task difficulty using RP80 as the dependent measure, and for 71 percent of the *R*-squared variance for quantitative task difficulty using Level as the dependent measure (see Tables 13).

Insert Table 11 about here.

In addition to accounting for a significant amount of *R*-squared variance for task RP80 and Level, the variables and constructs identified in this paper illustrate the internal consistency of processing characteristics within Levels. As shown in Tables 11, 12, and 13, we found that, based on approximately 50 RP80-point intervals, different combinations of variable scores appeared with consistent regularity within Levels. For instance in Level 1, prose tasks with the variable combination scores of 1, 1, and 2 or less occurred 85 percent of the time. In Level 2, prose tasks with the combination scores of 2, 2, and 2 (or 3, 3 or less, and 3 or less) appeared 98 percent of the time. And in Level 3, prose tasks with the combination scores of 4, 3 or less, and 3 or less occurred 79 percent of the time.

Insert Table 12 about here.

Table 11

Characteristics of Variables by Prose Task Difficulty Level for DOL and NALS Combined

	Level1 (RP80 >225)	Level 2 (RP80 225-275)	Level 3 (RP80 276-325)	Level 4 (RP80 326-375)	Level5 (RP80 <375)	*p* value for Level
Number of Tasks by Level (n= 71)	6	17	24	16	8	
Percentage of U.S. Population Completing Tasks at this Level	21%	27%	32%	17%	3%	
Means for RP80 Values	197	253	296	350	419	
Means for Readability	6.67	7.41	8.04	8.87	9.00	.77
Means for Type of Match	1.00	2.71	3.58	4.94	5.38	.00**
Means for Plausibility of Distractors	1.17	1.94	2.46	3.44	3.75	.00**
Means for Type of Information	1.83	2.53	2.88	3.50	4.25	.00**
Combination Scores	1, 1, 2 or less	2, 2, 2; or 3, 3 or less, 3 or less	4, 3 or less, 3 or less	4, 4 or higher, 4 or less	5, 5 or higher, 5 or less	R^2=80%
Percentage of Tasks within Level with Combination Scores	83% (5/6)	71% (12/17)	79% (19/24)	69% (11/16)	88% (7/8)	Overall 76% (54/71)

BEST COPY AVAILABLE

Table 12

Characteristics of Variables by Document Task Difficulty Level for DOL and NALS Combined

	Level1 (RP80 >225)	Level 2 (RP80 225-275)	Level 3 (RP80 276-325)	Level 4 (RP80 326-375)	Level5 (RP80 <375)	*p* value for Level
Number of Tasks by Level (n= 117)	26	43	28	10	108	
Percentage of U.S. Population Completing Tasks at this Level	23%	28%	31%	15%	3%	
Means for RP80 Values	198	250	301	345	400	
Means for Readability	2.81	4.58	5.64	6.30	7.50	.75
Means for Type of Match	1.42	2.47	3.50	4.00	5.90	.00**
Means for Plausibility of Distractors	1.73	2.28	3.11	4.30	4.20	.00**
Means for Type of Information	1.46	1.74	2.14	2.40	3.60	.00**
Combination Scores	1, 2 or less, 2 or less	2, 2, 2; or 3, 3 or less, 3 or less	4 or higher, 3 or less, 3 or less	4, 4 or higher, 4 or less	5, 5 or higher, 5 or less	R^2=87%
Percentage of Tasks within Level with Combination Scores	85% (22/26)	98% (42/43)	89% (25/28)	70% (7/10)	90% (9/10)	Overall 90% (105/117)

BEST COPY AVAILABLE

Overall, the internal consistency of prose tasks in the combined DOL and NALS assessments had an internal consistency within Level of 76 percent (see Table 11). The internal consistency of document tasks in the combined DOL and NALS had an internal consistency within Level of 90 percent (see Table 12). Finally, the internal consistency of quantitative tasks in the combined DOL and NALS had an internal consistency within Level of 83 percent (see Table 13).

Insert Table 13 about here.

This internal consistency of variable scores within Levels suggests that prose, document, and quantitative task difficulty builds upon consistent patterns of constructs. As noted in Table 14, these patterns of constructs include different combinations of strategy requirements which represent a scaffolding of task difficulty. For instance, we find most document tasks at Level 1 tend to require readers to identify information which is quite concrete, e.g., it represents a 'person,' 'thing,' 'amount,' 'type of,' 'temporal,' 'action,' or 'location.' Moreover, to complete such tasks, readers must locate a single piece of information which is identical to, or synonymous with, the information given in the question or directive. In most cases, there are no distractors at this Level, or if they do appear, they represent either given or new information but not both.

Like tasks in Level 1, many tasks in Level 2 ask readers to complete information which is fairly concrete. However, in Level 2, we find some tasks which also require readers to identify information representing 'manner,' 'goal,' 'purpose,' 'attempt,' 'alternative,' and 'condition' information. In addition, tasks at Level 2 often require readers to make a low-level inference, or identify a condition or an antecedent in order to identify requested information in a text. Finally, tasks at this Level tend to have a

Table 13

Characteristics of Variables by Task Difficulty Level for Quantitative Literacy Scales for DOL and NALS Combined

	Level1 (RP80 >225)	Level 2 (RP80 225-275)	Level 3 (RP80 276-325)	Level 4 (RP80 326-375)	Level5 (RP80 <375)	*p* value for Level
Number of Tasks by Level (n=71)	2	5	31	24	9	
Percentage of Population Completing Tasks at this Level	22%	25%	31%	17%	4%	
Means for RP80 Values	206	255	298	351	404	
Means for Readability	2.00	3.40	3.58	3.38	4.57	.19
Means for Type of Match	1.00	2.00	2.48	2.58	2.78	.57
Means for Plausibility of Distractors	1.00	1.60	2.45	2.83	3.67	.00**
Means for Type of Calculation	1.00	2.20	2.32	3.42	4.44	.00**
Mean Score Value for Operation Specificity	1.00	2.80	3.77	4.79	5.78	.00**
						R^2=71%
Combination Scores	1, 1, 2 or less	2 or 3, 2, 2 or less; or 3, 1, 1	4 or higher, 3 or less, 3 or less	4 or higher, 4, 4 or less; or 6, 4 or less, 4 or less	5, 5 or higher, 5 or less	
						Overall
Percentage of Tasks within Level with Combination Scores	100% (2/2)	80% (4/5)	84% (26/31)	79% (19/24)	89% (8/9)	83% (59/71)

BEST COPY AVAILABLE

distractor for either given or new information present but not in the same paragraph as the answer.

Tasks in Level 3 tend to require readers to identify condition information. In other instances, tasks require readers to identify a 'reason' or 'explanation.' In terms of type of match, Level 3 tasks again require readers to make literal, synonymous, and low-level inference matches between the question or directive and the text. Unlike Level 1 and 2 locate tasks, Level 3 tasks usually require readers to identify and list multiple responses (the number of which is specified in the question or directive). Also the questions and directives of Level 3 tasks tend to consist of several phrases. Moreover, these tasks generally require readers to complete requested information by identifying special conditional information stated in a question or directive or by establishing antecedence between a pronoun and its reference. Distracting information for both given and requested information tends to be present, both of which appear in different paragraphs from one another and neither of which appear in the same paragraph as the answer.

Insert Table 14 about here.

Table 14 presents the remaining constructs for document as well as prose and quantitative literacy for each of the five Levels of task difficulty.

Clothes Anthropometry and Adult Proficiency Profiles Reconsidered

A second purpose of this paper was to describe a new approach for "profiling" adult literacy proficiencies so that they can be equated as precisely as possible with task difficulty much as human size can be equated with clothes size. First, in implementing the profile approach, we began by determining the number of levels necessary to describe the full range of proficiencies in our sample. Hence, similar to the six size levels identified in Table 15 (i.e., 'X-small,' 'small,' 'medium,' 'large,' 'X-large,' and

Table 14

Constructs of Prose, Document, and Quantitative Literacy by Level of Task Difficulty

	Prose	Document	Quantitative
Level 1 0-225	Most of the tasks in this Level require readers to identify information which is quite concrete, including a 'person,' 'place,' or 'thing,' as well as an "attribute,' 'amount,' 'type of,' 'temporal,' 'action,' 'procedure,' or 'location.' Moreover, to complete these tasks, readers must process relatively short text to locate a single piece of information which is identical to (or synonymous with) the information given in the question or directive. If distractors appear in the text, they tend be located in a paragraph other than in the one in which the correct answer occurs.	Most of the tasks in this Level require readers to identify information which is quite concrete, including a 'person,' or 'thing,' as well as an 'amount,' 'type of,' 'temporal,' 'action,' or 'location.' Moreover, to complete these tasks, readers must process relatively brief documents to locate a single piece of information which is identical to (or synonymous with) the information given in the question or directive. In some cases, readers must enter personal information (e.g., their name and age) onto a document. If distractors appear in the document, they tend to be representative of either given or new information but not both.	Tasks in this Level require readers to perform a single, relatively simple arithmetic operation, which usually is addition. The numbers to be operated on tend to be adjacent to one another, appearing in the same row or column. The label(s) associated with these numbers are provided and the numbers to be operated on usually can be identified with little or no search. In most instances, the type of operation is signalled by an arithmetic sign (e.g., '+' or '-') or the term 'total' (when used to mean 'addition'). The numbers to be operated on are unrelated to previous tasks and require no tranformations. In most cases, there are no other amounts listed which could serve as plausible distractors.
Level 2 226-275	Like tasks in Level 1, many tasks in Level 2 ask readers to complete information which is fairly concrete. However, in Level 2, we find some tasks which also require readers to identify information representing 'manner,' 'goal,' 'purpose,' 'attempt,' alternative,' and 'condition' information. Moreover, tasks at Level 2 often require readers to make a low-level inference, or identify a condition or an antecedent in order to identify requested information in a text. Finally, tasks at this Level tend to have a distractor for either given or new information present but not in the same paragraph as the answer.	Like tasks in Level 1, most tasks in Level 2 ask readers to complete information which is quite concrete. However, in Level 2, we find some tasks which also require readers to identify 'condition' information. Moreover, tasks at Level 2 often require readers to make a two-feature match or a low-level inference to relate given information to information in a document. Other tasks require readers to make two or more dependent cycle matches between a legend and a graph, or between two documents. In other instances, tasks may require readers to integrate information within a document. Finally, tasks at this Level tend to have a distractor for both given and new information present but not in the same node as the answer.	Like tasks in Level 1, some tasks in Level 2 require readers to perform a simple addition. However, in Level 2, we also find tasks which require readers to perform a simple subtraction. As with Level 1 tasks, the numbers to be operated on in Level 2 tasks tend to be adjacent to one another, appearing in the same row or column. The label(s) associated with these numbers are provided. However, the numbers to be operated on usually require some search involving simple cycling. In some instances, operations are no longer signaled by arithemetic signs but rather in terms of relational statements such as 'how much more' and 'how much less.'. The numbers to be operated on again are unrelated to previous tasks and require no tranformations. However, in some instances, there are other numbers present which serve as simple distractors.

BEST COPY AVAILABLE

Level 3 276-325	Tasks in this Level tend to require readers to identify condition information. In other instances, task require readers to identify a 'reason' or 'explanation.' In terms of type of match, Level 3 tasks again require readers to make literal, synonymous, and low-level inference matches between the question or directive and the text. Unlike Level 1 and 2 locate tasks, Level 3 tasks usually require readers to identify and list multiple responses (the number of which is specified in the question or directive). Also the questions and directives of Level 3 tasks tend to consist of several phrases. Moreover, these tasks generally require readers to complete requested information by identifying special conditional information stated in a question or directive or by establishing antecedence between a pronoun and its reference. Distracting information for both given and requested information tends to be present, both of which appear in different paragraphs from one another and neither of which appear in the same paragraph as the answer.	Tasks in this Level often require readers to identify condition information in addition to the usual amount, temporal, and location information. In terms of type of match, Level 3 tasks again require readers to make literal, synonymous, and low-level inference matches between the question or directive and the document. However unlike Level 1 and 2 locate tasks, Level 3 tasks may require readers to compare or contrast information as well as identify simple patterns or trends . Also the questions and directives of Level 3 tasks tend to involve multiple feature matching involving tables which contain nested information. Distracting information for both given and requested information tends to be present, both of which may appear in the same node other than the node in which the answer occurs.	Besides requiring readers to perform simple addition and subtraction, tasks at Level 3 begin to include multiplication and division, as well as simple multiple operations (e.g., an addition followed by a subtraction). Unlike in Level 1 and 2 tasks, the numbers to be operated on in Level 3 tasks tend not to be adjacent to one another, although they continue to appear in the same row or column. Again as in Level 1 and 2 tasks, the label(s) associated with these numbers in Level 3 tasks are provided and, as in Level 2 tasks, the numbers to be operated on usually require some search invovling cycling. However, unlike Level 1 and 2 tasks, Level 3 tasks usually require readers to infer operations based on accounting terms (e.g., 'how much saved'). The numbers to be operated on again are unrelated to previous tasks. However, in some instances, tasks may require transformations of times and ratios. In most instances, there are other numbers present which serve as plausible distractors.

Level 4 326-375	Tasks in this Level tend to require readers to identify rather abstract information, including 'reason,' 'evidence,' 'explanation,' 'causation,' 'result,' 'comparison,' and 'contrast.' In terms of type of match, Level 4 tasks generally require readers not only to locate but also to cycle and integrate. Again, multiple responses may be required but for which the number of responses is not specified. As with Level 3 tasks, Level 4 tasks often require readers to complete requested information by identifying special conditional information stated in a question or directive, or by establishing antecedence between a pronoun and its reference. In other cases, high text-based inferences must be made to distinguish the correct requested information from distracting information. At this Level, distracting information for both given and requested information tends to be present, both of which may appear in the same paragraph as the answer.	Tasks in this Level sometimes require readers to identify rather abstract information, including 'contrast' and 'equivalence.' Level 4 tasks tend to require readers to make more difficult contrasts and to identify more complex patterns or trends than are characteristic of Level 3 tasks. Also, Level 4 tasks tend to involve multiple feature matching. At the same time, Level 4 tasks invariably include plausible distractors for both given and requested information which appear together in the same node which, in some cases, may be the answer node.	Most tasks at Level 4 require readers to perform calculations using multiplication and division, as well as more complex multiple operations (e.g., a multiplication followed by a subtraction). Numbers to be operated on in Level 4 tasks tend not to be adjacent to one another nor do they appear in the same row or column. In some instances, the label(s) associated with these numbers are not provided. In other instances, the labels for numbers are provided but readers must infer equations and operations based upon ratios inferred from statements in the document. In some instances, the numbers to be operated on are related to previous tasks. Moreover, some Level 4 tasks, like Level 3 tasks, may require transformations of times and ratios. And also like Level 3 tasks, there usually are other numbers present which serve as distractors. Some of these distractors may involve the situation where distractors for both given and requested information appear in the same node but in a node other than the answer.
Level 5 376-500	Tasks in this Level tend to require readers to identify quite abstract information, including 'contrast,' 'equivalence,' and 'theme' (or 'summary'). In terms of type of match, Level 5 tasks often require readers not only to locate, cycle, and integrate but also generate. Generate may involve the use of specialized background knowledge to interpret a phrase or to synthesize text information. At this Level, distracting information for both given and requested information may be present, both of which frequently appear in the same paragraph as the answer.	Tasks in this Level tend to require readers to identify rather abstract information, including 'contrast' and 'patterns.' Level 5 tasks again tend to involve multiple feature matching but also require greater degrees of inferencing in matching information and in identifying requested information. Some of these tasks require readers to provide multiple responses but do not designate how many responses are needed. These tasks also require readers to identify conditional information in identifying requested information. Concomitantly, Level 4 tasks often include plausible distractors for both given and requested information which appear in the same node as the answer.	All tasks at Level 5 require readers to perform calculations involving multiple operations (which usually include multiplication and division). Numbers to be operated on in Level 5 tasks tend not to be adjacent to one another nor do they appear in the same row or column. In some instances, the label(s) associated with these numbers are not provided. In other instances, the labels for numbers are provided but readers must infer equations and operations based upon known ratios provided in the document. Moreover, some Level 5 tasks, like Level 3 and 4 tasks, may require transformations of times and ratios. Finally, there are other numbers usually present which serve as distractors. Some of these distractors may appear in the same node as the answer.

'XX-large') which anthropometrists have used to describe the range of shirt sizes found within an adult male population, we have identified five levels (i.e., Levels 1 through 5) which describe the range of literacy proficiencies found within the U. S. adult population (ages 16 and older).

Second, in implementing the profile approach, we identified a set of variables which significantly account for the range of observable differences within our population using Level as the predictor variable. Hence, similar to the variables of neck, chest, and waist size, arm length, and height which account for size variance of men's torsos, we identified type of information, type of match, plausibility of distractors, type of calculation, and operation specificity as variables which significantly account for literacy proficiency variance within the U.S. adult population.

Third, in implementing the profile approach, we have identified a range of variable scores and their concomitant constructs which enable us to interpret the nature of variance within and between proficiency levels. Hence, similar to the anthropometrists' specifications that size XX-large represents a neck size between 18 and 18½ inches, a chest size between 50 and 52 inches, a waist size between 44 and 46 inches, and a regular arm length between 38 and 38½ inches (see Table 15), we have noted that Level 5 prose proficiency consists of the values of 5, 5 or higher, and 5 or higher for type of information, type of match, and plausibility of distractors. Moreover, we have interpreted what these values mean in terms of their strategy requirements as described in Table 14; these strategy descriptions are comparable to the numbers in Table 15 which provide interpretation of torso sizes and their underlying variables.

Insert Table 15 about here.

Fourth, in implementing the profile approach, we have established the probabilities which predict how well adults with a known RP80 proficiency score are

Table 15

Mens' Body Part Measurements by Shirt Sizes

	X-Small		Small		Medium		Large		X-Large		XX-Large	
Neck	13	13 1/2	14	14 1/2	15	15 1/2	16	16 1/2	17	17 1/2	18	18 1/2
Chest	28	32	34	36	38	40	42	44	46	48	50	52
Waist	24	26	28	30	32	34	36	38	40	42	44	46
Arm (Reg.)[1]	31 1/2	32	32 1/2	33	33 1/2	34	34 1/2	35	35 1/2	36	36 1/2	37
Arm (Tall)[2]	—	—	34	34 1/2	35	35 1/2	36	36 1/2	37	37 1/2	38	38 1/2

[1]Regular = 5'8"-6'0" [2]Tall = 6'1"-6'3"

likely to perform tasks of varying difficulty. Hence, similar to anthropometrists who have determined the probability of fit between male torso size and shirt size (Boorstin, 1973), we have determined the probability of fit between adult proficiency and task difficulty. For instance, as shown in Table 1, we can predict, knowing an adults' RP80 proficiency score, the probability by which these adults would be able to process individual tasks. Moreover, based on Figure 1, we can further predict the likelihood with which adults would be able to process tasks within a certain Level of difficulty. Hence, as Figure 1 suggests, an adult with a proficiency score of 250 would have approximately a 90 percent chance of performing Level 1 document tasks, an 80 percent chance of performing Level 2 tasks, a 50 percent chance of performing Level 3 tasks, a 20 percent chance of performing Level 4 tasks, and a ten percent chance of performing Level 5 tasks.

Fifth, based on such known probabilities, we are able, in implementing the profile approach, to determine the optimal type of match between an individual's RP80 score and a task's Level score. In anthropometry, this fit often is represented by a range of probabilities between men's torso size and shirt size which requires the least amount of tailoring to accommodate shirt fit to torso (Boorstin, 1973). In educational measurement and instruction, what constitutes the optimal fit may be viewed differently, depending upon what our intent is (Nitko, 1989)—e.g., to determine an individual's functional, instructional, or frustration reading level, or the likelihood of workers being able to adequately perform in a given occupational domain (Painchaud & Jezak, 1994). Discrepancies between literacy proficiency and performance fits may be defined in terms of the length and cost of instruction needed to bring an individual to a criterion level of proficiency (Mikulecky & Drew, 1991).

Based on the information which the profile approach yields for matching literacy tasks to adult readers, we can begin to reinterpret the notion of standards in terms of increased precision resulting from successively more complex levels of computer-based

measurement (Bunderson et al., 1989). This third purpose of our paper is addressed in the following section.

Defining U. S. Literacy Standards in Terms of DOL/NALS' Five Levels of Proficiency and Bunderson et al.'s (1988) Four Generations of Measurement

In his work, *The Americans: The democratic experience*, Boorstin (1973) argues that progress in any discipline is usually possible only when a standard definition is advanced and uniformly interpreted by a community or society. Notes Boorstin, this is particularly true in the area of educational measurement. However, while our society has come to accept most standard definitions found in the physical and biological sciences, we have not been so quick to develop broad consensus of what constitutes standard operational definitions of performance and proficiency and their underlying constructs in the social sciences (Mosenthal & Kamil, 1991). In short, despite all the research and reviews on the topic of setting proficiency and performance standards (Andrew & Hecht, 1976; Behuniak, Archambault, & Gable, 1982; Beuk, 1984; Glass, 1978; Hambleton & Eignor, 1980; Jaeger, 1989; Koffler, 1980; Pearson, 1993; Popham, 1978; Purves, 1993; Shepard, 1980, 1984), such standards continue to be set unsystematically in adult literacy.

While various norm-and criterion-referenced approaches have been advanced for setting standards in adult literacy (Kirsch & Jungeblut, 1986; Mosenthal & Kirsch, 1989a; Stedman & Kaestle, 1987), both have significant limitations. Perhaps the principal limitation is the lack of interpretability—or the ability to understand what an adult's placement along a scale means when compared to the relative proficiencies of other adults, and when compared to the relative difficulty of tasks which comprise a performance domain (Bunderson et al., 1989; Kirsch et al., 19893; Kirsch & Guthrie, 1980; Mosenthal & Kirsch, 1989a). While norm-referenced approaches (e.g., Nafziger et al., 1975) enable comparisons to be made between adult literacy proficiencies as measured in terms of grade levels, such approaches provide little

understanding of what it means to be reading at, say, a fourth versus a sixth grade reading level. Similarly, while criterion-referenced approaches (e.g., Harris et al., 1970; Murphy, 1973; Northcutt, 1975) enable us to identify the percentages of adults who are able to perform one or more tasks at a particular criterion level, such approaches provide little or no understanding of the constructs which underlie this performance across an entire task domain (Mosenthal & Kirsch, 1989a).

In contrast to the traditional norm-and criterion-referenced approaches to setting standards in adult literacy, we have described in this paper a profile approach which addresses these limitations of interpretability. As developed to date, this approach enables us to interpret what it means for adults to be proficient in performing task representing different Levels of difficulty in the domains of prose, document, and quantitative literacy. More importantly, this approach has the potential for significantly increasing the precision of fit in matching adult learner proficiencies with task difficulty, as has been suggested by Bunderson et al. (1989) in their discussion of computerized educational measurement.

As a first step for improving the standardization of fit between adult proficiency and task difficulty, the profile approach could be adapted to computerized testing by converting present paper-and-pencil tasks to a *conventional computer-administered format*. In this generation of measurement, task difficulty parameters will have to be recalibrated to determine the effects (if any) of this new format. The advantage of using the computer testing mode is that, because the DOL and NALS tasks have open-ended scoring, the computer could be programmed to identify the variations of acceptable and nonacceptable responses, thus greatly enhancing the speed and reliability in scoring and reporting test results.

In a more advanced generation of measurement, the profile approach could be used to design computer-administered tests in which the presentation of each successive task is *computer adaptive*. Adaptive here means that the selection and presentation of

successive tasks depends upon adults' responses on earlier tasks. This could be accomplished as follows. First, because research (Kirsch et al., 1992, 1993) has determined which adult background characteristics correlate highest with literacy Level, these background characteristics could be used to determine the initial starting (or reference) point for administering items (e.g., since Hispanic adults with 9 to 12 years of education are most likely to have a mean prose literacy proficiency around 200, (Kirsch et al., 1994), tasks with difficulty values in this RP80 range would first be administered to adults with these background characteristics).

Next, based on calibrated task response probabilities (such as those shown in Table 1), three or more tasks might be selected and presented with the likelihood of being completed correctly with an estimated probability. Depending upon how well adults respond at, say, the 80 percent probability level, new tasks would then be selected and presented which represented successively higher or lower probabilities of being completed correctly. Once adults' zones of proximal proficiency have been determined (e.g., where adults perform tasks correctly, say, 80 to 90 percent of the time), successively more difficult tasks would be administered to establish a "probability proficiency profile" specific for each adult. Such a profile would predict adults' ability to perform different types of tasks at different Levels of difficulty in the areas of prose, document, and quantitative literacy.

Concomitant with this method of tailoring tasks to adults' proficiencies, a second means for selecting and presenting tasks would be to make adaptation decisions based on construct difficulty. For instance, if adults demonstrate that they are able to successfully complete tasks which involve identifying a thing using a two-feature locate match with no distractors, the next task might involve identifying a manner using a three-feature locate match with distractors for given and new information in different nodes but in a node other than in which the answer appears. Based on adults' patterns of performance on tasks representing different types of constructs, this would enable the

computer to establish a "construct proficiency profile" for each adult. Such a profile would provide a diagnosis of the strategies that adults were adept and not adept at employing in completing literacy tasks.

The advantage, of course, of using computer adaptive testing to determine a standardized fit between an adults' proficiency and task difficulty is that more tasks could be administered which yield more precise diagnostic information about adults' proficiencies (Nitko, 1989). Such computer adaptive testing stands in contrast to the conventional tests administered by computer or paper which tend to have high measurement precision near the average test score but which have low measurement precision for adults with low and high proficiencies (Bunderson et al., 1989; Hambleton, 1989). In contrast, the use of a computer adaptive test (such as the one describe here) could yield a much higher level of measurement precision for all adults due to the ability of the computer to tailor task selection and presentation to responses representing the full range of adult proficiencies.

In a third generation of measurement, the profile approach could be used to design *continuous measurement systems.* Such systems use calibrated measures embedded in a curriculum to continuously and unobtrusively estimate the dynamic changes in adult learners' changing proficiencies. Changes might be reported in terms of adults' increased probabilities for performing more difficult tasks, their increased probabilities for performing successfully within a task difficulty Level, or their improved ability to employ more sophisticated processing strategies in the context of structurally more complex stimuli.

The distinguishing characteristic of continuous measurement is the ability to specify dynamically adult learners' position in the "growth space" (i.e., Levels 1-5) of the prose, document, and quantitative scales (cf. Bunderson et al., 1989). By identifying adults' changing probability and construct proficiency profiles over the course of instruction, a "trajectory of learning" could be identified. This trajectory

could be used to establish how adults progress not only between Levels but also within Levels as well. The point here, of course, is that, while many adults may have the same RP80 and Level proficiency scores when initially tested, they may have these scores for very different reasons (Hambleton, 1989). One advantage of using continuous measurement is to determine specifically those strategies which adults are adept and not adept at learning both within and between Levels and then to dynamically tailor instruction accordingly.

As Bunderson et al. (1989, p. 387) have noted, continuous measurement assumes a two-part definition of curriculum: "(a) a course of experiences laid out to help the learner grow toward certain educational ends, that is, a path through a domain; (b) a set of course markers, or standards, that serve as milestones of accomplishment along the way, that is, beginning, intermediate, and terminal markers." The profile approach provides for both. In terms of (a), the profile approach identifies five Levels of constructs which represent different scaffolds of prose, document, and quantitative strategies. Within each Level, there are different combinations of strategies and processing demands associated with type of information, type of match, plausibility of distractors, type of calculation, and operation specificity. Taken together, these strategies and their concomitant processing demands define the highly functional domains of prose, document, and quantitative literacy.

As such, knowledge of these strategies, variables, Levels, and domains could easily provide the basis for designing a continuous-measurement curriculum for improving adults' abilities to function in today's society and workplace (Heath, 1980; Kirsch et al., 1993; Kirsch et al., 1994; Painchaud & Jezak). Once such a curriculum had been established in such a way that the constructs taught were the same as the constructs tested (cf. Hambleton, 1989; Nitko, 1989), this would then enable teachers to systematically determine the learning trajectories that adults used to become increasingly proficient in the domains of prose, document, quantitative literacy. This

could be accomplished by having the computer identify the prototypical performance patterns associated with adults' continuously updated probability and construct proficiency profiles.

In terms of (b), the profile approach would provide functional-literacy curriculum designers with knowledge of how to design instructional and measurement tasks so that these tasks reflected different combinations of construct characteristics. Once their difficulty parameters had been determined and validated using RP80 scaling, these tasks would automatically represent "standards of accomplishments" at a micro level in terms of: (a) their known probabilities of being performed as individual tasks relative adults' currently identified proficiency scores, (b) their known probabilities of being performed as tasks representative of a given difficulty Level relative to adults' currently identified proficiency score; (c) their known probabilities of being performed as tasks representing a particular combination of constructs relative to adults' performance on tasks reflecting similar strategies. Moreover, these tasks (once calibrated in terms of their RP80 values and construct characteristics relative to other tasks on general prose, document, and quantitative scales) would further represent "standards of accomplishments" at a macro level in terms of the overall five Levels of literacy with Level 1 perhaps serving as the "beginning accomplishment marker," Level 5 as the "terminal marker," and Levels 2, 3, and 4 representing a range of "intermediate markers."

In a fourth generation of measurement, the profile approach could possibly be used as a first step towards designing *intelligent instructional systems* Such systems attempt to simulate the expert knowledge associated with a domain (Bunderson et al., 1989). Furthermore, a system is said to be intelligent to the extent that it can quickly reconfigure curriculum to continuously and unobtrusively estimate the dynamic changes in adult learners' changing proficiencies.

To be useful instructionally, such systems will have to represent several experts' knowledge. First, such a system will have to represent the knowledge of experts who perform within a domain; this knowledge will consist of the experts' understanding of the full-range of task contents, strategies, and structures which define highly functional performance within this domain (Bunderson et al., 1989). Second, such a system will have to represent the knowledge of expert teachers who are able to efficiently and effectively assess adults' learning trajectory patterns so that optimal feedback and instruction in a task domain's contents, strategies and structures may be provided. And third, such a system will have to represent the knowledge of response interpretation experts who have knowledge not only of the various ways of interpreting the quality of learner responses, but who also have the ability to interpret routine patterns of errors (or "buggy routines") (Nitko, 1989).

Besides being programmed with these different types of expert knowledge, such intelligent instructional systems will also have to be equipped with different levels of "novice knowledge" that adult learners are likely to possess. Equipped with such knowledge, intelligent instructional systems would then have the capability of providing appropriate feedback when queried by novices who may want to clarify their own understanding (or lack thereof) about a variety of concerns, which may range from a desire to better understand a strategy or how to pose a question to the computer. Ideally, such a system would further enable learners to add the results of their queries back into the knowledge base, thereby enabling the system to provide a richer personalized system for answering future queries specific to different individuals. This process of querying the system and then using information gleaned would constitute a new performance domain for which proficiency scales would have to be constructed and their underlying constructs identified.

At this point, such systems remain futuristic considerations rather than immediate realizations. However, in the anticipation of such systems, the profile

approach could serve as important paradigm for operationalizing learners' proficiencies in querying intelligent instructional systems. Moreover, this approach could perhaps be further used to characterize the proficiencies of expert teachers, response interpretation experts, and domain performance experts in the domain of general and workplace literacy tasks found in the twenty-first century.

Summary

At one level, this paper has served to describe and validate the constructs underlying adults' prose, document, and quantitative literacy proficiencies as measured using the DOL and NALS assessments. Since much adult literacy research, policy, and practice is currently being based on the findings of these assessments at the national, state, and local levels (Mosenthal, in press; Wagner, Tuijman, & Kirsch, in press), this paper has attempted to provide an interpretation of what it means to be proficient at each of the five Levels on the prose, document, and quantitative literacy scales. As Mosenthal (in press) and others (e.g., Kirsch & Guthrie, 1980) have argued, without such knowledge of the constructs which underlie performance and proficiency on assessments, researchers, policy makers, and practitioners have little or no basis for understanding what constitutes the problems of adult literacy; consequently, they have no basis for knowing how to set reasonable literacy goals for adults. Nor do these groups have an informed basis for deciding: (a) which groups should benefit from nationally and state supported literacy programs, (b) what actions should be taken to solve adult literacy problems and achieve adult literacy goals, and (c) what should characterize the nature of evaluation to ensure that adult literacy programs are achieving what they were designed to achieve.

In addition to these concerns, this paper has further attempted to define and illustrate a new method for conceptualizing standards of proficiency and performance in the domain of adult literacy. In contrast to other metaphors for setting standards (cf. Pearson, 1993), we have argued that defining adult literacy proficiency is not unlike

the anthropometry problem of fitting clothes to humans—only in adult literacy the problem is finding the best fit between adults' proficiency and task difficulty rather than finding the best fit between adult torso size and shirt size. To provide a means for standardizing the fit between proficiency and task difficulty so that this fit can be made efficiently and effectively for a broad range of adults in a cost-effective manner, we have proposed a "profile approach" for accomplishing this. As we have noted, this approach has been made possible largely from the benefit of several national adult-literacy assessments (Kirsch & Jungeblut, 1986; Kirsch et al., 1993; Kirsch et al., 1994) which have involved the administration of a large number of tasks to a national population of adults (ages 16 and older),and which have used item response theory to scale tasks and proficiencies.

Although we have acknowledge the utility of the anthropometry metaphor as a way of understanding how "fit" can be standardized, we have also suggested that this metaphor tends to be based on static rather than on dynamic (or continuous) measurement. While this may be generally adequate for the clothes industry where the growth trajectories of adults are quite slow (and recalibration of torso sizes for individuals is infrequently required), this is much less adequate in the case of educational programs where the growth trajectories of adult learners can be dramatic (and the recalibration of individual proficiency is constantly required) (Mikulecky & Drew, 1991). To address this need for recalibration, we have attempted to illustrate, using Bunderson et al.'s (1989) four generations of measurement, how a standardized set of proficiency Levels, with their accompanying interpretive constructs, could be used to continuously profile the changing growth patterns of adults and, concomitantly, could be used in decisions of how to tailor instruction so that each adult's literacy proficiency may be optimally enhanced.

References

Andrew, B. J., & Hecht, J. T. (1976). A preliminary investigation of two procedures for setting examination standards. *Educational and Psychological Measurement, 36*, 35-50.

Behuniak, P., Jr., Archambault, F. X., & Gable, R. K. (1982). Angoff and Nedelsky standard setting procedures: Implications for the validity of proficiency test score interpretation. *Educational and Psychological Measurement, 42*, 247-255.

Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research, 56*, 137-172.

Bassett, E. D., Goodman, D. G., & Fosegan, J. S. (1981). *Business records controls* (5th ed.). Cincinnati, OH: South-Western Publishing.

Beuk, C. H. (1984). A method for researching a compromise between absolute and relative standards in examinations. *Journal of Educational Measurement, 21*, 147-152.

Boorstin, D. J. (1973). *The Americans: The democratic experience.* New York: Random House.

Bunderson, C. V., Inouye, D. K., & Olsen, J. B. (1989). The four generations of computerized educational measurement. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 367-407). New York: Macmillan.

Burch, J. G., & Grudnitski, G. (1986). *Information systems: Theory and practice* (4th ed.). New York: Wiley.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse processes: Advances in research and theory* (Vol. 1) (pp. 1-40). Norwood, NJ: Ablex.

Cross, D., & Paris, S. (1987). Assessment of reading comprehension: Matching test purposes and test properties. *Educational Psychologist, 22*, 331-339.

Fry, E. (1977). Fry's readability graph: Clarifications, validity, and extension to level 17. *Journal of Reading, 21*, 242-252.

Glass, G. V. (1978). Standards and criteria. *Journal of Educational Measurement, 15*, 237-261.

Guthrie, J. T., Seifert, M., & Kirsch, I. S. (1986). Effects of education, occupation, and setting on reading practices. *American Educational Research Journal, 79*, 220-227.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 147-200). New York: Macmillan.

Hambleton, R., K., & Eignor, D. R. (1980). Competency test development, validation, and standard setting. In R. M. Jaeger & C. K. Tittle (Eds.), *Minimum competency testing: Motives, models, measures, and consequences* (pp. 367-396). Berkeley, CA: McCutchan.

Harris, L., & Associates. (1970). *the 1971 national reading difficulty index: A study of functional reading ability in the United States, for the National Reading Center.* Washington, DC: The National Reading Center.

Heath, S. B. (1980). The functions and uses of literacy. *Journal of Communications, 30*, 123-133.

Hartley, J. (1985). *Designing instructional text.* New York: Nichols.

Hempel, C. G. (1966). *Philosophy of natural science.* Englewood Cliffs, NJ: Prentice Hall.

Jaeger, R. M. (1989). Certification of student competence. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 485-514). New York: Macmillan.

Kibby, M. W. (1993). What reading teachers should know about reading proficiency in the U. S. *Journal of Reading, 37*, 28-40.

Kirsch, I. S., & Guthrie, J. T. (1980). Construct validity of functional reading tests. *Journal of Educational Measurement, 17*, 81-93.

Kirsch, I. S., & Guthrie, J. T. (1984). Adult reading practices for work and leisure. *Adult Education Quarterly, 13*, 485-507.

Kirsch, I. S., & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults* (NAEP Report No. 16-PL-01). Princeton, NJ: Educational Testing Service.

Kirsch, I. S., & Jungeblut, A. (1992). *Profiling the literacy proficiencies of JTPA and ES/UI populations: Final report to the Department of Labor.* Princeton, NJ: Educational Testing Service.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey.* U. S. Department of Education: Washington, DC.

Kirsch, I. S., Jungeblut, A., & Mosenthal, P. B. (1994). *Moving toward the measurement of adult literacy.* Paper presented at the March NCES meeting, Washington, DC.

Kirsch, I. S., & Mosenthal, P. B. (1989). Building documents by combining simple lists. *Journal of Reading, 33*, 132-135.

Kirsch, I. S., & Mosenthal, P. B. (1990a). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly, 25*, 5-30.

Kirsch, I. S., & Mosenthal, P. B. (1990b). Nested lists. *Journal of Reading, 33*, 294-297.

Kirsch, I. S., & Mosenthal, P. B. (1992). How to navigate a document using locate known/need-to-know strategies. *Journal of Reading, 36*, 140-143.

Kirsch, I. S., & Mosenthal, P. B. (1992/1993). Integration strategies: Higher order thinking applied to documents. *Journal of Reading, 36*, 322-327.

Kirsch, I. S., & Mosenthal, P. B. (1991). Understanding definitions, descriptions, and comparison/contrasts. *Journal of Reading, 35,* 156-160.

Koffler, S. L. (1980). A comparison of approaches for setting proficiency standards. *Journal of Educational Measurement, 17,* 167-178.

Lands' End Direct Merchants. (1994, May). How to get the correct size. *Summer Catalogue.* Dodgeville, WI: Lands' End.

Linn, R. L. (Ed.)(1989). *Educational measurement* (3rd ed.). New York: Macmillan.

Meyer, B. J. F., & Marsiske, M., & Willis, S. L. (1993). Text processing variables predict the readability of everyday documents read by older adults. *Reading Research Quarterly, 28,* 235-248.

Mikulecky, L. (1982). Job literacy: The relationship between school preparation and workplace actuality. *Reading Research Quarterly, 7,* 400-419.

Mikulecky, L., & Drew, R. (1991). Basic literacy skills in the workplace. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2) (pp. 669-689). White Plains, NY: Longman.

Miller, R. B. (1984). Transaction and structures and format in form design. In R. Easterby & H. Zwaga (Eds.), *Information design: The design and evaluation of signs and evaluation of printed materials* (pp. 529-544). New York: Wiley.

Mosenthal, P. B. (in press). Using large-scale assessments to set educational policy: An agenda-analytic approach. In D. Wagner, A. Tuijman, & I. S. Kirsch (Eds.), *Adult basic skills: Innovations in the measurement and policy analysis.* New York: Hampton Press.

Mosenthal, P. B. (1976-77). Bridge principles in an abridged reply to Goodman. *Reading Research Quarterly, 12,* 586-603.

Mosenthal, P. B., & Kamil, M. (1991.). Understanding progress in reading research. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2) (pp. 1013-1046). White Plains, NY: Longman.

Mosenthal, P. B., & Kirsch, I. S. (1989a). Designing effective adult literacy programs. *Poetics, 18,* 239-256.

Mosenthal, P. B., & Kirsch, I. S. (1989b). Lists: The building blocks of documents. *Journal of Reading, 33,* 58-60.

Mosenthal, P. B., & Kirsch, I. S. (1989c). Intersecting lists. *Journal of Reading, 33,* 210-213.

Mosenthal, P. B., & Kirsch, I. S. (1991a). Information types in nonmimetic documents: A review of Biddle's wipe-clean slate. *Journal of Reading, 34,* 654-660.

Mosenthal, P. B., & Kirsch, I. S. (1991b). Extending prose comprehension through knowledge modeling. *Journal of Reading, 35,* 58-61.

Mosenthal, P. B., & Kirsch, I. S. (1991c). Toward an explanatory model of document literacy. *Discourse Processes, 14,* 147-180.

Mosenthal, P. B. & Kirsch, I. S. (1992). Cycle strategies in document search: From here to there to wherever. *Journal of Reading, 36,* 238-242.

Mosenthal, P. B., & Kirsch, I. S. (1993a). Generate strategies: Coping without cues and clues. *Journal of Reading, 36,* 416-419.

Mosenthal, P. B., & Kirsch, I. S. (1993b). Profiling students' document strategy abilities. *Journal of Reading, 36,* 578-583.

Mosenthal, P. B., & Kirsch, I. S. (1993c). Profiling students' quantitative literacy abilities: An approach with precision. *Journal of Reading, 36,* 668-674.

Mosenthal, P. B., & Kirsch., I. S. (in press). What makes document processing difficult: A profile approach. *Reading Research Quarterly.*

Murphy, R. T. (1973). *Adult functional reading study* (Final report, Project No. 09004, PR-73-48). Princeton, NJ: Educational Testing Service.

Nafziger, D. H., Thompson, R. B., Hiscox, M. D., & Owen, T. R. (1975). *Tests of functional adult literacy: An evaluation of currently available instruments.* Portland, OR: Northwest Regional Educational Laboratory.

National Assessment of Educational Progress. (1972). *Reading: Summary* (Report 02-R-00). Denver, CO: Education Commission of the States.

National Assessment of Educational Progress. (1976). *Functional literacy: Basic reading performance.* Denver, CO: Education Commission of the States.

Nitko, A. J. (1989). Designing tests that are integrated with instruction. In R. L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 447-474). New York: Macmillan.

Northcutt, N. (1975). *Adult functional competence: A report to the Office of Education Dissemination Review Panel.* Austin, TX: University of Texas.

Painchaud, G.,. & Jezak, M. (1994, April). *Language use in the workplace: Entry-level jobs in the service industry.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Pearson, D. P. (1993). Standards for the English language arts: A policy perspective. *Journal of Reading Behavior, 25,* 457-475.

Popham. (1978). *Setting performance standards.* Los Angeles: Instructional Objectives Exchange.

Purves, A. C. (1993). Setting standards in the language arts and literature classroom and the implications for portfolio assessment. *Educational Assessment, 1,* 175-200.

Rayner, D. (1982). *Review of administrative forms report to the Prime Minister.* London, England: Management & Personnel Office.

Ryan, D. E. (1880). *Human proportions in growth: Being the complete measurements of the human body for every age and size during the years of juvenile growth.*

Shepard, L. A. (1980). Standard setting issues and methods. *Applied Psychological Measurement, 4,* 447-467.

Shepard, L. A. (1984). Setting performance standards. In R. A. Berk (Ed.), *A guide to criterion-referenced test construction* (pp. 169-198). Baltimore, MD: Johns Hopkins University Press.

Sticht, T. G. (1978). *Literacy and vocational competency.* (Occasional Paper No. 39, National Center for Research in Vocational Education). Columbus, OH: Ohio State University.

Stedman, L. C., & Kaestle, C. F. (1987). Literacy and reading performance in the United States, from 1880 to the present. *Reading Research Quarterly, 22,* 8-46.

Venezky, R. L. (1982). The origins of the present-day chasm between adult literacy needs and school literacy instruction. *Visible Language, 16,* 113-127.

Wagner, D., Tuijman, A., & Kirsch, I. S. (Eds.) (in press). *Adult basic skills: Innovations in the measurement and policy analysis.* New York: Hampton Press.